The Dissertation Committee for Sebastián Souyris
certifies that this is the approved version of the following dissertation:

Models to Predict and Influence Consumer Demand: Applications to Television Advertising and Solar Panel Adoption

Committee:

Anant Balakrishnan, Supervisor

Jason Duan, Co-Supervisor

Guoming Lai

Varun Rai

Sridhar Seshadri

Stathis Tompaidis

Models to Predict and Influence Consumer Demand: Applications to Television Advertising and Solar Panel Adoption

by

Sebastián Souyris

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

August 2019

Dedicada al Juaco.

Acknowledgments

There are many people to whom I am deeply grateful for their support, advice, work, and friendship to conclude this project. Many thanks to Sridhar Seshadri for his ongoing support since we met at NYU 10 years ago. Countless thanks to my supervisors, Anant Balakrishnan, and Jason Duan. Without their dedication, ideas, patience, and work, this project would not be finished. Thanks to Varun Rai for opening the door to the energy research space by supporting me with conferences, data, and deep insight into solar. Thanks to Guoming Lai and Stathis Tompaidis for having been members of my committee. Thanks to Steve Gilbert and Leon Lasdon for their classes and guidance. Thanks to Sriram Subramanian for discussing ideas, support, and working on the TV ads scheduling project. Thanks to Shiv Sehgal and Mukesh Sehgal for facilitating the ads project. I am grateful to many colleagues for the solid foundation, opportunities, and friendship. Many thanks to Andrés Weintraub, Cristián Cortés, Fernando Ordóñez, Rafael Epstein, René Caldentey, Denis Sauré, Mike Pinedo, Nikolay Osadchiy, Gustavo Vulcano, Qi Wu, Dana Popescu, Andrea Prado, Martín Quinteros, José Guajardo, Marcel Goic, Antoine Sauré, and Jaime Miranda. Thanks to my supportive and fellow UT Ph.D. students, Vivek Vasudeva, Chinmoy Mohapatra, Gonzalo Maturana, Carlos Parra, Jacelli Cespedes, Alvin Leung, Juan Andrade, Abhishek Roy, and Magdalena Saldaña. A heartfelt thanks to my friends Matías Bulnes, Pietro González, and Pablo Casals. They pushed me to continue when I needed it most. Thanks to my family Pamela, Keka, Daniela, Angélica, and all the Souyris. Thanks to my Dad, Raúl, who gave me life, an education, friendship, and everything in between. Last but not least, many thanks to my awesome wife, Kelli Stretesky, who has supported me unconditionally during these previous two years and especially during the sprint of finishing the dissertation. Thank you, my Love.

Models to Predict and Influence Consumer Demand: Applications to Television Advertising and Solar Panel Adoption

Publication No. ____________________

Sebastián Souyris, Ph.D.
The University of Texas at Austin, 2019

Supervisors: Anant Balakrishnan
Jason Duan

This dissertation consists of two independent essays that share the common goal of predicting and influencing consumer demand: (i) Scheduling Advertising on Television; and (ii) Peer Effects in the Diffusion of Solar Panels: A Dynamic Discrete Choice Approach. These essays are data-driven problems of operations and management science, which we address using optimization and econometrics.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This dissertation consists of two independent essays that have the same ultimate goal: to influence consumers to make decisions that are aligned with the objectives of a business or policy maker. In the following two sections, we introduce the real-world context, goals, analytical methods, and main contributions of each essay.

1.1 Scheduling Advertising on Television

A major source of revenue for any cable television network is the placement of advertisements, or *spots*, within its programming. The contracts between the advertisers and the network require that the network delivers an expected number of views from a targeted demographic during the course of the advertising campaign. This is called the *target viewership*. These contracts also specify the number of *spot-seconds* used to deliver the required number of impression. The network must deliver the contracted number of spot-seconds even if the target viewership has already been achieved. Advertisers also require the network to deliver the target demographic viewership according to an agreed-upon pace during the course of the campaign.

A contract between the advertiser and the network is referred to as a *deal*. In a deal, the advertiser also places *constraints* on scheduling spots, which could include requirements that spots must air on a specified day of the week or at a specified time of day. Constraints can also require that spots are shown during specific programs or requires that spots air in the first or last position of the commercial break. On the other hand, the network typically sets aside a certain amount of time every hour for airing spots. The network also aims to maintain the right balance between content and spots so as to retain audience engagement.

The finite supply of air time available in which the network can air spots is referred to as the available *inventory*. From the network's perspective, in order to make the most effective use of their limited inventory, spots must be scheduled within a time slot that meets the following goals: The first goal of the network is to deliver viewership from the target demographic requested by the advertiser, and delivers this viewership at the required pace. This viewership must not exceed the number of impressions requested by the advertiser, because impressions beyond the contracted quantity do not generate additional revenue. The second goal is to minimize the viewership from any demographic outside the targeted demographic, since these impressions are not monetized.

The projections for national viewership accepted by networks and advertisers are those issued by Nielsen Holdings N.V., referred to as *Nielsen ratings*. These ratings are published a few days after programs are broadcast. Therefore, one sub-task of a network is to forecast viewership based on ratings data from the past and schedule spots using the forecast so the two goals defined above are met. The quality of the schedule obviously dependents on the quality of the forecast.

In Chapter 2, we develop a theoretically based framework for solving the spots scheduling problem described above. This scheduling is achieved in three stages. We start in Stage 1 with a multi-day optimization model that takes the aggregate demand for spots as input to generate weights for the deals in play. These weights indicate the relative importance of deals after taking into account the number of impressions that remain to be delivered for the deal during the remaining lifetime of the campaign. These weights are fed into in a single-day optimization model that assigns spots to breaks so as to make the best use of the viewership, given the existing inventory and the available spots. This optimization is performed to deliver a higher share of impressions to deals that have larger weights.

For computational efficiency, we split the single-day optimization model that schedules spots in breaks into two stages. In Stage 2, the spots are allocated to the breaks. Then, the precise positioning of the allocated spots within each break is completed in Stage 3. To forecast ratings, we developed a time series model that takes advantage of the correlation between the ratings of breaks to protect against uncertainty. In Chapter 2, we provide structural results that justify the use of a point forecast instead of a more sophisticated approach, such as stochastic programming or robust optimization. The framework is packaged in a decision support system that is used by leading television networks in the United States and in India, and it has produced an increase in revenue of 3% to 5%, which translates to approximately $60 million annually for one of the larger users.

1.2 Peer Effects In The Diffusion Of Solar Panels: A Dynamic Discrete Choice Approach

Installations of solar photovoltaic (PV) systems in the United States have overgrown in the last few years, mainly credited to federal, state, and local level incentives. The federal government established a 30% federal investment tax credit (ITC) in 2006 that runs until this year, 2019, and then it is scheduled to decline. State support for PV diffusion is also in a downturn. The most ambitious state policies, the California Solar Initiative, is already in its last phase and only available through municipal service providers. Simultaneously, lower electricity prices in several parts of the United States due to low natural gas prices have further pressured the continued penetration of PV.

On the other hand, PV installation costs have come down dramatically: Average residential prices were under $5 per watt in 2013 compared to $10 per watt in 2007. How do these price and incentive dynamics impact the diffusion of PV? In particular, amidst these shifts, how should electric utilities rethink their solar programs under budgetary constraints? A better understanding of these issues may be used to design solar programs that are tailored to target specific segments of the population.

To address these questions, we studied the Austin, Texas, residential PV market in Chapter 3 leveraging a rich data set that includes each household demographics, property characteristics, and electricity consumption. We joined that data with the adopters PV installation period, installation cost, rebate received, and system attributes. We developed a dynamic discrete choice (DDC) model that permits the exploration of the effects of policy and market shifts on the diffusion of PV. Accurately, our model characterizes the factors in PV diffusion at a consumer disaggregated level modeling households as consumers who are looking forward: At every period, each household who has not adopted yet compares the decision of adopting today versus waiting and deciding in the next period. In dynamic programming terminology, we formulated the household decision of adoption as an optimal stopping problem.

The factors that enter into consideration in the utility function of each household are the net present value of installing a PV (which includes the cost of installation, the local rebate, the ITC, and the savings from PV electricity generation) plus the awareness of the technology that the household gained from neighbors who installed a PV previously, the peer effects. For estimation of the structural parameters that govern behavior, we designed and implemented a Bayesian Markov Chain Monte Carlo method that allows us to handle heterogeneity in a dynamic context. Because the model is structural—in the sense that the parameters capture the preferences of the households and quantifies the spatial peer effects—we can project the dynamics of the market and do counterfactual analysis, thereby, giving insights about what rebate schedules are more efficient for accelerating the diffusion process.

The value of the estimated parameters shows that wealthy households are more keen to install a PV and that the peer effects are significant and play an essential role in the decision of adoption. Moreover, the marginal effect of a new neighbor adopter is higher for households that are rural, then suburban and finally urban. Furthermore, unobserved consumer heterogeneity is considerable as measured by the standard deviation of the random effects that capture the heterogeneity.

For validation, we benchmarked our model against the classic Bass model comparing the predicted number of adopters per quarter in a two-year test set. Our DDC model outperform the Bass model on MAPE by order of magnitude on the test set. For counterfactual policy analysis, we assumed the hypothetical situation that the government could adjust the rebate every quarter to match a pre-announced net cost of installation. We created the following net cost scenarios over time: observed net cost as the base case; constant net cost; stepwise decreasing net cost; and stepwise increasing net cost. Subsequently, we predicted the number of adopters and computed the budget spent on rebates under each scenario. Constant and decreasing net cost of installation are Pareto superior to the base case in the sense that these scenarios incentivize more adopters at a lower total budget spent than the base case. An increasing net cost incentivizes many more adopters than the base case but at higher total budget spent. By using our framework, a policymaker can measure what would be the effect of different rebate schedules and adjust it according to her objective.

Chapter 2

Scheduling Advertising on Television

2.1 Introduction

A major source of revenue for any cable television network is the placement of advertisements, or *spots*, within its programming. The contracts between the advertisers and the network require that the network delivers a target viewership—specifically, a number of impressions of a particular demographic during the course of the advertising campaign. The demographics that advertisers seek are defined according to two dimensions: gender and age group. For example, *M18-34* refers to males between the ages of 18 and 34, *F18-34* refers to females between the ages of 18 and 34, and *P18-34* refers to persons (males and female) between the ages of 18 and 34. The contracts also specify the number of spot-seconds used to deliver the required number of impressions. The contracted number of spot-seconds must be delivered even if the target viewership has already been achieved. Advertisers also require the network to deliver the target demographic viewership according to an agreed upon pace during the course of the campaign. The contract between the advertiser and the network is referred to as a *deal*. The particular type of deal described here is referred to as *guaranteed* deal because the number of spots must be telecast and the target viewership must be achieved as promised. In addition, the advertiser places *constraints* on scheduling spots, which could include requirements that spots must air at a specified day of the week and time of day, requirements that spots are shown during specific programs, or requirements that spots air in the first or last position of the commercial break. These guaranteed deals comprise most of the revenue for a typical North American cable network. There are also deals that are not guaranteed. For these types of deals, the network is allowed to decide whether to schedule the spots, and no viewership targets are required. In this case, the advertiser pays the network on a per–telecast spot basis.

The network typically sets aside a certain amount of time every hour for airing spots. The network also aims to maintain the right balance between content and spots so as to retain audience engagement. The finite supply of air time available for the network to air spots is referred to as the available *inventory*. From the network's perspective, in order to make the most effective use of their limited amount of inventory, spots must be scheduled within a time slot so that the following goals are met: (i) The viewership is delivered in the target demographic requested by the corresponding advertiser at the required pace, which is the rate at which spots must be scheduled (e.g., the number of spots per day). In addition, the viewership must not exceed the number of impressions requested by the advertiser, because impressions beyond the contracted quantity do not generate additional revenue. And, (ii) the viewership that is delivered in any demographic outside the target must be minimal, as these impressions are also not monetized.

The inventory is further segmented into short time intervals, or *commercial breaks* (15 seconds to few minutes), that are distributed across many programs scheduled to air. The viewership for the break usually depends on the program containing the break, the time of day, and the day of the week. The projections for national viewership accepted by networks and advertisers are those issued by Nielsen Holdings N.V., referred to as Nielsen ratings. These ratings are published a few days after programs are broadcast. Therefore, one sub-task is to forecast the viewership based on ratings data from the past and schedule spots using the forecast so that the goals defined above are met. The quality of the schedule is obviously dependent on the quality of the forecast.

For each day, the qualifying spots are scheduled to air within the breaks to produce a *log*, or a 24–hour (one-day) schedule with programs interspersed among the breaks that contain spots. To schedule these effectively, the decision-maker must consider several other factors in addition to the scheduling constraints requested by the advertiser. At any given point in time, each deal has already delivered some of the target impressions and has a specific number of spots to run, a number of impressions to deliver, and a number of days remaining until completion. Some deals may exceed the required pace (i.e., they are over-performing) and some may fall below the required pace (i.e., they are under-performing). The network's ability to correct any mismatch will depend on the time remaining on the deal. The advertiser pays only for the contracted number of impressions: Any impressions beyond this number would be a service that the network has given away, and any shortfall of impressions requires the network to pay a penalty, either in the form of additional spots or in cash. Therefore, it is important for the network to ensure that the deals track as closely as possible to the required number of impressions.

In this essay, we developed a framework that is theoretically grounded for solving the spots scheduling problem described above. The approach is bundled in an advertisement revenue management support system, that is used by leading television networks in the United States and India producing increase in revenue of 3% to 5%, that translate in about $60M annually for one of the bigger users. The scheduling is achieved in three stages. We start in Stage 1 with a multi-day optimization model that takes the aggregate demand for spots as input to generate weights for the deals in play. These weights indicate the relative importance of deals after taking into account the number of impressions still due to be delivered for the deal during the remaining lifetime of the campaign. The more the deal is lacking in the required pacing (i.e., the more significant amount of impressions pending relative to the time left on the deal) the higher is the weight of the deal. These weights fed into in a single-day optimization model that assigns spots to breaks so as to make the best use of the viewership, given the existing inventory and the available spots. This is done so as to deliver a higher share of the impressions to the deals with larger weights. For computational efficiency, we split the single-day optimization model that schedules spots in breaks in two stages. In Stage 2, the spots are allocated to the breaks. Then, the precise positioning of the allocated spots within each break is completed in Stage 3. This approach generates a schedule for one day at a time, preferably at the latest possible point in time so that the layout of programs and commercial breaks (i.e., time and durations) as well as the set of spots available to be scheduled are as close as possible to the final version. Delaying the scheduling until the latest possible point in time has the additional benefit of generating the best possible forecast for viewership ratings. These models are supported by a ratings forecasting model that takes advantage of the correlation between ratings of breaks to protect against uncertainty. We implemented classic time series techniques to forecast ratings, and provide structural results that justify the use of a point forecast instead of a more sophisticated approach, such as stochastic programming or robust optimization.

We focus on describing the solution approach and the mathematical modeling. This description includes several modeling ideas and features that can guide similar efforts. We also touch upon several practical and business considerations and describe how these are managed in practice. As with any implementation, one measure of success is the potential for increase in revenue. In this regard, this application, which is being used daily at leading networks, provides an increase in revenue of the magnitude of tens of millions dollars annually. Another measure of success is whether the application has led to more problems for analysis and solution. In that respect, several new problems are identified and are currently being solved to improve both the quality of the solutions as well as the analysis of the results for use in the ad-sales process.

The organization of the essay is as follows: Section 2.2 describes the related academic literature. Section 2.3 gives an overview of the problem and the solution approach and presents our structural results that supports the use of ratings point forecast. Section 2.4 details the optimization models. Section 2.5 specifies the solution strategy. Section 2.6 describes the statistical models used to forecast the ratings. Section 3.6 presents numerical results, and Section 3.8 concludes.

2.2 Literature Review

Advertising in television is a multi-billion dollar industry in which revenue management (RM) techniques can have a considerable effect on the bottom line. Despite this, as [75] detail, only a few RM studies in the television industry had been documented in the operations literature until 2005. However, in recent years there has been a growing interest; see survey by Pandey et al. [50]. For a full description of the television industry, see Blumenthal and Goodenough [9]. The cable network business processes can be grouped into four sub-processes: Planning of creative goods (see, for example, Caves and Guo [19]), Scheduling shows (Headen et al. [37], Reddy et al. [58], Rust et al. [63], Wilbur et al. [84], etc.), Inventory sales and Spots scheduling. Our work is most closely related to the last two as discussed below.

The sales of audience inventory occurs in two periods (Phillips and Young [52]), the *upfront market* and the *scatter market* (*forward* and *spot* markets in finance parlance). The upfront market takes place approximately four months before the broadcast season, when the networks publicize the shows schedule. At this time, the advertisers buy ratings points that are guaranteed. All the remaining inventory that is unsold during the upfront market (about 20%–40%), is sold on the scatter market during the broadcasting season. To sell inventory during the upfront market, it is necessary to have projections of the ratings inventory and the advertisers demand. Bollapragada et al. [12] developed a decision support system based on integer programming and heuristics to support the sales process at the NBC's television network. Subsequent work has been reported on advertiser demand forecasting (Bollapragada et al. [14]) and how much inventory to sell on each market period (Bollapragada and Mallik [11], Kimms and Muller-Bungart [44], Zhang [85]). Araman and Popescu [4] define the allocation problem of stochastic ratings between upfront and scatter markets as the *media revenue management* or capacity planning problem, and formulate it as a random yield multiple lot-sizing in a production to order system (Grosfeld-Nir and Gerchak [36]); the authors obtain structural results that define the optimal capacity allocation depending on the contract parameters, audience, and time. Carbajal and Chaar [18] implemented an integer programming approach to solve this problem at a major U.S. television network. Using a different approach, Banciu et al. [6] examine network bundling strategies to sell to advertisers. These bundles can be composed of different demographic audiences or air times. In this essay, take the sales as given and only concern ourselves with the day to day scheduling to fulfill the sale commitments. Once the content (i.e., the programming) is acquired and scheduled, and after the deals have been signed for spots, the spots must be scheduled at the right times to make the most effective use of air time. This essay deals with this process. The first paper on this topic is Bollapragada et al. [13], who solve the scheduling problem of programming a given set of commercials of the same duration that should be aired a pre-specified number of times as uniformly as possible with respect to break position. The authors formulate the model as a network flow problem with a non-linear loss objective function, and they construct ad-hoc heuristics that can produce good solutions for test instances. For the same problem and test instances, Brusco [16] improves the solution time and optimality gap using a specialized branch and bound procedure and a simulated annealing heuristic, and also tests alternative loss functions. Brusco and Singh [17] incorporate time separability conditions and the possibility of having spots of different durations. Bollapragada and Garbiras [10] developed heuristic methods for scheduling spots in breaks using an integer programming model with real world conditions such as advertisers preferences for specific positions within the break. This system is used by NBC. Gaur et al. [30] extend the previous model by adding flexibility to model commercial conflicts and by developing a specialized algorithm. Zhang [85], as a second step, proposes a quadratic integer program to minimize the deviations from the original schedule generated in the previous process of inventory sales. Their formulation is a simplification of the real world problem and it is separated by shows, which results in one small problem per show that is easy to solve. This essay differs from the previous cited literature in that we developed a detailed spots schedule to minimize the penalty of under delivery, while honoring a myriad of constraints. Our work does not separate per show; we can move spots from one show to another. In fact, several constraints apply for the entire day (and some across days); therefore, the scheduling problem is complex. A related problem is scheduling spots for live television, in which the length of the breaks are uncertain, so the decision to air spots must be made in real time (Popescu and Crama [53]). The problem that we deal with in this essay is within cable television, where the length of the breaks are known at the moment of scheduling.

In the following section, we present the mathematical model, solution approach and structural results that support the use of point forecast ratings as input for the optimization.

2.3 Problem and Description of Solution Approach

The television spots scheduling problem can be mathematically formulated as a multi-stage stochastic programming model where the uncertainty is on future ratings. The time horizon is $t = 1, \dots, T$ days and there is a countable finite set $\mathcal{S}$ of scenarios. The probability of scenario $s \in \mathcal{S}$ is p^s, such that $\sum_s p^s = 1$. The scenarios are connected in a tree where the consecutive stages of the tree are the days and at the last stage (i.e., stage T) there are $|\mathcal{S}|$ leaves; each leaf corresponding to one scenario. Furthermore, in each stage t there are $\mathcal{N}(t)$ set of nodes; such that, for each node $n \in \mathcal{N}(t)$, the set $\Omega(n) \subseteq \mathcal{S}$ are the scenarios that pass through node n. Because this is a tree, every scenario belongs to one and only one node at each stage (i.e., $\Omega(n) \cap \Omega(n') = \emptyset$, for all $n \neq n' \in \mathcal{N}(t)$, and $\bigcup_{n \in \mathcal{N}(t)} \Omega(n) = \mathcal{S}$ for all $t = 1, \dots T$). For modeling purposes, the root node of the tree is at stage 0, so that in stage $t = 1$ there are $\mathcal{N}(1)$ nodes. Specific to spots scheduling problem, the set of breaks that will air per day, $\mathcal{B}$, and the set of deals, $\mathcal{D}$, are provided. The viewership is segmented in the set

$\mathcal{Q}$ of demographics. For notional convenience through the essay, let indexes $b \in \mathcal{B}$, $d \in \mathcal{D}$ and $q \in \mathcal{Q}$. The time length of break b is L_b, and the length of a spot of deal d is H_d. Each deal d has a target demographic. Let the binary parameter G_{dq} be equal to 1 if deal d targets demo q and 0 if not. Furthermore, each deal d has a contracted number of impressions to be delivered, I_d. The cost per thousand impressions of demographic q is CPM_q. The rating of demo q in break b of day t in scenario s is r_{qb}^{st}. Let us define $CPM_d := \sum_{q \in \mathcal{Q}} G_{dq} CPM_q$, and $r_{db}^{st} := \sum_{q \in \mathcal{Q}} G_{dq} r_{qb}^{st}$, for all $d \in \mathcal{D}; b \in \mathcal{B}; t = 1, \ldots, T$ and $s \in \mathcal{S}$. The decisions variables are the following: x_{db}^{st}, the number of spots of deal d to schedule in break b of day t in scenario s; and y_d^s, the total number of shortfall impression of deal d in scenario s. The optimization problem is to minimize the expected shortfall:

$$
\begin{aligned}
\text{minimize} \quad & \sum_{s \in \mathcal{S}} \sum_{d \in \mathcal{D}} p^s y_d^s CPM_d \\
\text{subject to} \quad & \sum_{d \in \mathcal{D}} H_d x_{db}^{st} \leq L_b && \forall\, s \in \mathcal{S}, b \in \mathcal{B}, t = 1, \ldots, T, && (2.1) \\
& \sum_{t=1}^{T} \sum_{b \in \mathcal{B}} r_{db}^{ts} x_{db}^{st} + y_d^s \geq I_d && \forall\, s \in \mathcal{S}, d \in \mathcal{D}, && (2.2) \\
& \boldsymbol{x}^{ts} \in A(t) && \forall\, s \in \mathcal{S}, t = 1, \ldots T, && (2.3) \\
& \boldsymbol{x}^{ts} = \boldsymbol{x}^{ts'} && \forall\, s \neq s' \in \Omega(n), n \in \mathcal{N}_t, t = 1, \ldots, T, && (2.4) \\
& x_{db}^{ts} \in \mathbb{Z}_{\geq 0}, && \forall\, s \in \mathcal{S}, d \in \mathcal{D}, b \in \mathcal{B}, t = 1, \ldots, T, && (2.5) \\
& y_d^s \in \mathbb{Z}_{\geq 0}, && \forall\, s \in \mathcal{S}, d \in \mathcal{D}. && (2.6)
\end{aligned}
$$

Constraint (2.1) restricts the maximum slots-time allocated to each break; constraint (2.2) ensures that each deal gets assigned the contracted number of impressions minus the shortfall. Constraint (2.3) restricts the schedule in each day t to the set $A(t)$, that are linear restrictions which can theoretically restrict to schedulable assignments (specific constraints are provided in Section 2.4). Constraint (2.4) are the nonanticipativity constraints. And constraints 2.5 and 2.6 are the integer non negative nature of variables.

This general problem of assigning spots to exact positions is very challenging computationally and fraught with practical difficulties of incomplete information and uncertainty. In any given day multiple deals compete for spots in the schedule. Some deals have been running for a long time and have either been under-performing or over-performing. Some deals are due to end soon (i.e., next week) whereas some will not expire for some time (i.e., this quarter). Furthermore, the creative work for the spots are not ready until the day prior to airing and therefore their relative lengths, specific requirements as well as their restrictions are unknown for future schedules. Finally, the ratings of programs are not known. Forecasts of ratings are imprecise and become less precise with increasing time lengths.

2.3.1 Solution Approach

Solving the large program is impossible given today's state of the art machines and software. Instead we chose to focus on the day-to-day scheduling problem. Even that problem is broken into three stages due to running time considerations. Figure 2.1 shows the schematic processes and information flow. The sequence adopted is as follows:

Stage 1: Estimate Deal Weights. Every week, the model is solved to determine the relative weights for deals. The inputs to this model are (i) all the contractual details of the deals for which spots are under consideration for inclusion in the coming week, (ii) how the deals have performed relative to their target audience, (iii) the number of weeks remaining in the deals, and (iv) information regarding spots yet to be scheduled for each deal. The output from this stage comprises the deal weights that are used in the next two stages to produce the daily schedule. The deal weight reflects the relative price per impression given to the deal in the desired demographic.

Stage 2: Schedule Spots in Breaks. The network keeps fine-tuning the details of the schedule showing the minute-by-minute airing of shows, advertisements and promotions, or logs, until nearly the evening prior to airing the schedule. These initial logs are created manually based on expert opinion and provide a feasible schedule that finally is optimized. The decision to continue the practice of manually scheduling the log is made based on several considerations. The management typically does not want the entire schedule to be managed by a machine. The schedulers prefer to decide which spots to air and whether to give favored slots to certain deals. Moreover, the log is developed several weeks prior to airing. In the initial versions, the log is quite incomplete and filled with generic information or placeholders that become more specific as the date of airing approaches. Therefore, it is practical to allow the schedulers to develop the log gradually and deliver it for optimization at the final step. The log is downloaded from the production system on to the optimization server. Several pre-checks are performed to determine data accuracy as well as automatic detection of rule violations by schedulers who may occasionally relax rules to satisfy a client requirement or make a human decision to override the constraint. These violations are frozen in the final schedule. The initial schedule as well as the final schedule produced by the optimization are vetted by a traffic system. This system has a scheduling engine which automatically checks most of the rules imposed by the network and, more importantly, checks the rules imposed by government regulators. Therefore, the initial feasibility is guaranteed unless the scheduler has revised the schedule and has overridden the scheduling engine. This “close-to-final” schedule is optimized in Stage 2 to assign spots to breaks.

Stage 3: Arrange Spots in Break Positions. This model arranges the spots inside the breaks to which they are assigned in the previous stage. To the user, Stages 2 and 3 happen simultaneously and are not visible as separate steps. Afterward, the optimized schedule is pushed back into the log of the network using integration software. The traffic system verifies the schedule suggested by the optimizer. The human scheduler may make further changes until he or she is satisfied and then pushes the schedule to the production system. The production system sends the schedule for broadcast. The spots are only one component of what is finally aired. The actual programs, the creative for the spots, the promotional advertisements for the network's own programs, the local TV station content, and the local advertisements are all assembled before reaching the subscriber.

The resulting schedule and revenue gains are scrutinized on a regular basis. The metrics of performance are both quantitative and qualitative. The evaluation frequencies are daily, weekly, and monthly. Some sample metrics are:

- The revenue difference between the original schedule and the optimized schedule helps measure the lift in advertisement revenue once the actual ratings are obtained. The accounting is straightforward and the details are omitted.

- Other metrics are used, such as the distribution of spots of an advertiser or an agency by time, the day-to-day performance of deals, the use of prime time spots, the analysis of the source of revenue lift, and the forecast performance.

The management of the performance of deals to monitor the quality of the schedule is termed *stewardship*. The stewardship team uses the above information to analyze the effects of deal positions and the effects of the number of spots released for production on deal performance. The knowledge gained is used to fine-tune Stages 1 and 2. For example, the network may add additional spots to improve the condition of the deal, optimization can increase weights on these deals to improve delivery, or more prime time spots can be reserved for these deals.

The three optimization stages use as input the point forecast of ratings. Alternatively, it is possible to include the ratings uncertainty in the optimization models by using, for example, stochastic programming or robust optimization. In the following section, we establish structural results that justify this choice (of point forecasts) and demonstrate that there is negligible revenue loss in the long term if certain mild conditions are met. Moreover, in Section 2.6.1, we discover structural properties of the correlation between break ratings that allow us to reduce the size of the scheduling problem considerably and deal with uncertainty.

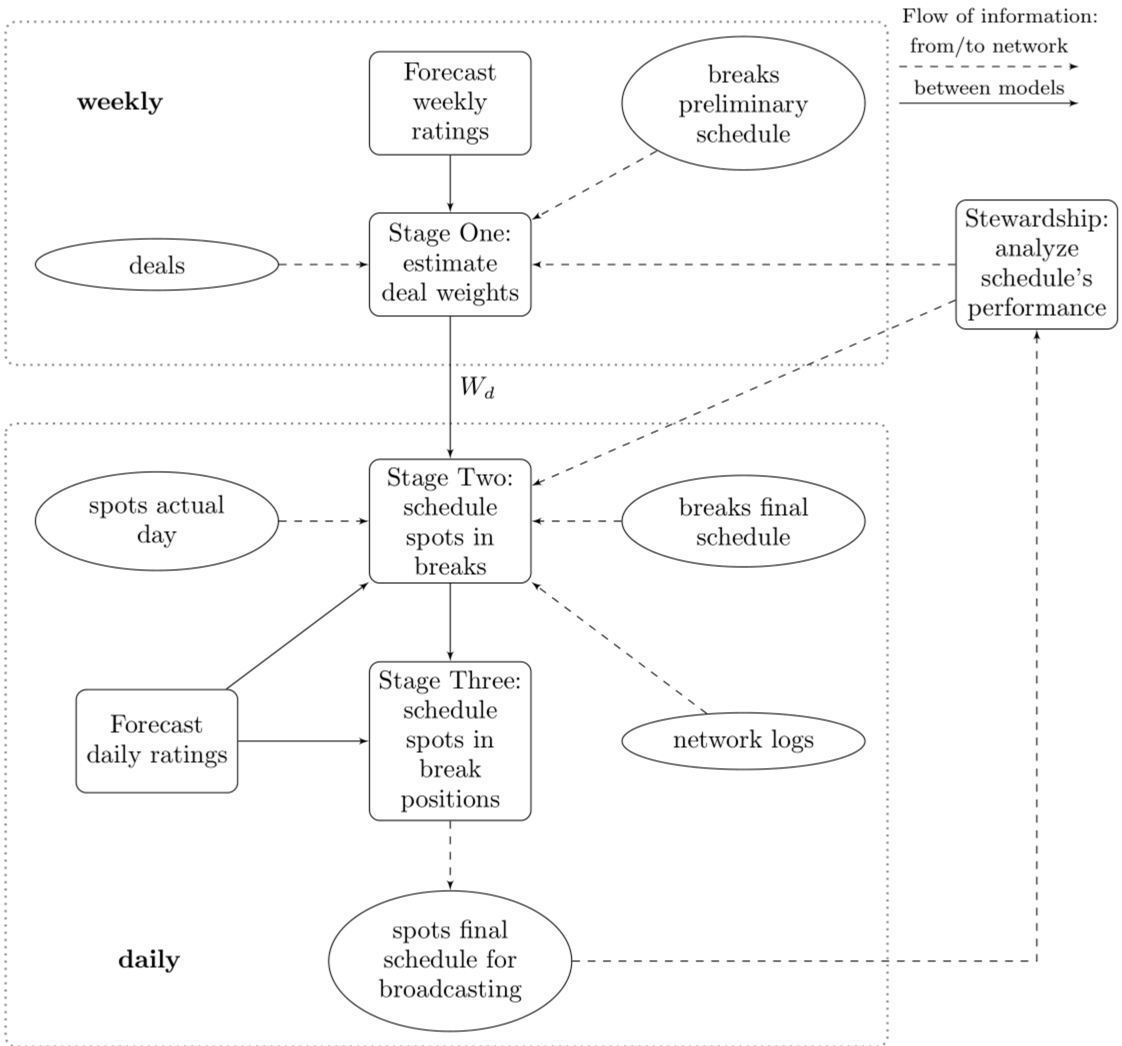


Figure 2.1: Processes and information flow.

2.3.2 Structural Results: Validity of Using Point Forecast

We start with a simplified version of the spot scheduling problem. The time horizon is one week, for which the set of breaks that will air, $\mathcal{B}$, and the set of deals, $\mathcal{D}$, are provided. The viewership is segmented in demographics, $\mathcal{Q}$. Each deal $d \in \mathcal{D}$ has a target demographic, which is identified by the binary parameter G_{dq}, for all $q \in \mathcal{Q}$. The parameter is equal to 1 if deal d targets demo q and 0 if not. Furthermore, each deal has a contracted number of impressions to be delivered, I_d. The cost per thousand impressions of demographic q is

CPM_q. The length of break b is L_b, and the length of a spot of deal d is H_d. The ratings in break b of demo q is a random variable, R_{qb}, which takes values in a set of scenarios $\mathcal{S}$ (we use scenarios for the construction of the proofs of Lemma 2.3.1, 2.3.2 and 2.3.3, which we further generalize for proof of Lemma 2.3.4). For each demo q in break b the ratings in scenario $s \in \mathcal{S}$ is r^s_{qb} with probability p^s. Let us define $CPM_d := \sum_{q\in\mathcal{Q}} G_{dq} CPM_q$, and $r^s_{db} := \sum_{q\in\mathcal{Q}} G_{dq} r^s_{qb}$, for all $d \in \mathcal{D}, b \in \mathcal{B}, s \in \mathcal{S}$. The decisions variables are the number of spots of deal d to schedule in break $b \in \mathcal{B}$, x_{db}, and the number of shortfall impression of deal d in scenario s, y^s_d. The optimization problem is to minimize the expected shortfall:

$$
\begin{aligned}
\text{(P)} \quad \text{minimize} \quad & \sum_{s\in\mathcal{S}} p^s \sum_{d\in\mathcal{D}} CPM_d y^s_d \\
\text{subject to} \quad & \sum_{d\in\mathcal{D}} H_d x_{db} \le L_b && \forall\, b \in \mathcal{B}, && (\mu_b) && (2.7)\\
& \sum_{b\in\mathcal{B}} r^s_{db} x_{db} + y^s_d \ge I_d && \forall\, s \in \mathcal{S}, d \in \mathcal{D}, && (\pi^s_d) && (2.8)\\
& x_{db} \ge 0 && \forall\, d \in \mathcal{D}, b \in \mathcal{B}, \\
& y^s_d \ge 0 && \forall\, s \in \mathcal{S}, d \in \mathcal{D}.
\end{aligned}
$$

The following lemmas allow us to estimate the impact of using the mean instead of the random ratings.

Lemma 2.3.1. *Replacing the random ratings r^s_{db} by their mean $\bar{r}_{qb}$ produces a lower bound of the value of the objective function of problem (P).*

To prove the lemma, notice that problem (P) is always feasible. Let μ_b and π^s_d be the dual variables corresponding to constraints (2.7) and (2.8) respectively. Also, let x^*_{db}, y^{s*}_d be the optimal solution of (P), and z_P^* be the optimal objective function. Its dual is

$$
\begin{aligned}
\text{(D)} \quad \text{maximize} \quad & \sum_{b \in \mathcal{B}} L_b \mu_b + \sum_{s \in \mathcal{S}} p^s \sum_{d \in \mathcal{D}} I_d \pi_d^s \\
\text{subject to} \quad & H_d \mu_b + \sum_{s \in \mathcal{S}} p^s r_{db}^s \pi_d^s \leq 0 && \forall\, d \in \mathcal{D}, b \in \mathcal{B}, && (x_{db}) && (2.9) \\
& \pi_d^s \leq CPM_d && \forall\, s \in \mathcal{S}, d \in \mathcal{D}, && (y_d^s) && (2.10) \\
& \mu_b \leq 0 && \forall\, b \in \mathcal{B}, \\
& \pi_d^s \geq 0 && \forall\, s \in \mathcal{S}, d \in \mathcal{D}.
\end{aligned}
$$

Let μ_b^*, π_d^{s*} be the optimal solution of problem (D), and z_D^* be the optimal objective function. By the Strong Duality Theorem, $z_P^* = z_D^*$. By the complementary slackness conditions, x_{db}^*, y_d^{s*} is optimal solution of the primal, and μ_b^*, π_d^{s*} optimal solution of the dual, if and only if

$$
\begin{aligned}
& \left(L_b - \sum_{d \in \mathcal{D}} H_d x_{db}^* \right) \mu_b^* = 0 && \forall\, b \in \mathcal{B}, \\
& \left(I_d - \sum_{b \in \mathcal{B}} r_{db}^s x_{db}^* - y_d^{s*} \right) \pi_d^{s*} = 0 && \forall\, s \in \mathcal{S}, d \in \mathcal{D}, \\
& \left(H_d \mu_b^* + \sum_{s \in \mathcal{S}} p^s r_{db}^s \pi_d^{s*} \right) x_{db}^* = 0 && \forall\, d \in \mathcal{D}, b \in \mathcal{B}, \\
& \left(\pi_d^{s*} - CPM_d \right) y_d^{*s} = 0 && \forall\, s \in \mathcal{S}, d \in \mathcal{D}.
\end{aligned}
$$

Using the mean of ratings ($\bar{r}_{qb} = \sum_{s \in \mathcal{S}} p^s r_{qb}^s$), problem (P) becomes:

$$
\begin{aligned}
(\bar{\text{P}}) \quad \text{minimize} \quad & \sum_{d \in \mathcal{D}} CPM_d \bar{y}_d \\
\text{subject to} \quad & \sum_{d \in \mathcal{D}} H_d \bar{x}_{db} \leq L_b && \forall\, b \in \mathcal{B}, && (\bar{\mu}_b) && (2.11) \\
& \sum_{b \in \mathcal{B}} \bar{r}_{db} \bar{x}_{db} + \bar{y}_d \geq I_d && \forall\, d \in \mathcal{D}, && (\bar{\pi}_d) && (2.12) \\
& \bar{y}_d, \bar{x}_{db} \geq 0 && \forall\, d \in \mathcal{D}, b \in \mathcal{B}.
\end{aligned}
$$

The decision variables under the mean of ratings are $\bar{x}_{db}$, number of deal d's spots to schedule in break b, for all $d \in \mathcal{D}$ and $b \in \mathcal{B}$; and $\bar{y}_{db}$, the number of shortfall impression of deal d, $d \in \mathcal{D}$. The corresponding dual variables of constraints (2.11) and (2.12) are $\bar{\pi}_d$ and $\bar{\mu}_b$.

Let $\bar{x}^*_{db}, \bar{y}^*_d$ be the optimal solution of problem ($\bar{\text{P}}$), and $\bar{z}^*_P$ be the optimal objective function. Its dual is:

$$
\begin{aligned}
(\bar{\text{D}}) \quad \text{maximize} \quad & \sum_{b\in\mathcal{B}} L_b\bar{\mu}_b + \sum_{d\in\mathcal{D}} I_d\bar{\pi}_d \\
\text{subject to} \quad & H_d\bar{\mu}_b + \bar{r}_{db}\bar{\pi}_d \le 0 && \forall\, d \in \mathcal{D}, b \in \mathcal{B}, && (\bar{x}_{db}) && (2.13) \\
& \bar{\pi}_d \le CPM_d && \forall\, d \in \mathcal{D}, && (\bar{y}_d) && (2.14) \\
& \bar{\mu}_b \le 0 && \forall\, b \in \mathcal{B}, \\
& \bar{\pi}_d \ge 0 && \forall\, d \in \mathcal{D}.
\end{aligned}
$$

Let $\bar{\mu}^*_b, \bar{\pi}^*_d$ be the optimal solution of ($\bar{\text{D}}$), and $\bar{z}^*_D$ be the optimal objective function. By the Strong Duality Theorem, $\bar{z}^*_P = \bar{z}^*_D$. By the complementary slackness conditions, $\bar{x}^*_{db}, \bar{y}^*_d$ is optimal solution of the primal $\bar{\text{P}}$, and $\bar{\mu}^*_b, \bar{\pi}^*_d$ optimal solution of the dual $\bar{\text{D}}$, if and only if

$$
\begin{aligned}
& \left(L_b - \sum_{d\in\mathcal{D}} H_d\bar{x}^*_{db}\right)\bar{\mu}^*_b = 0 && \forall\, b \in \mathcal{B}, \\
& \left(I_d - \sum_{b\in\mathcal{B}} \bar{r}_{db}\bar{x}^*_{db} - \bar{y}^*_d\right)\bar{\pi}^*_d = 0 && \forall\, d \in \mathcal{D}, \\
& \left(H_d\bar{\mu}^*_b + \bar{r}_{db}\bar{\pi}^*_d\right)\bar{x}^*_{db} = 0 && \forall\, d \in \mathcal{D}, b \in \mathcal{B}, \\
& \left(\bar{\pi}^*_d - CPM_d\right)\bar{y}^*_d = 0 && \forall\, d \in \mathcal{D}.
\end{aligned}
$$

Using the previous results, the proof of Lemma 2.3.1 is straightforward:

Proof of Lemma 2.3.1. The optimal solution of problem ($\bar{\text{D}}$) is feasible for problem (D). Therefore, $\bar{z}^*_D \le z^*_D$, and then $\bar{z}^*_P \le z^*_P$. $\square$

Lemma 2.3.2. *If there is a shortfall in every scenario for every deal, then the solution found in Lemma 2.3.1 is optimal to problem (P).*

Proof. A shortfall in every scenario implies that $y_d^s > 0$, for all $s \in \mathcal{S}, d \in \mathcal{D}$, for every feasible solution, in particular the optimal solution $y_d^{s*} > 0$. Then, because of the complementary slackness conditions, $\pi_d^{s*} = CPM_d$, for all $s \in \mathcal{S}, d \in \mathcal{D}$, and $\mu_b^* = -\max_{d \in \mathcal{D}} \left\{ \frac{CPM_d}{H_d} \sum_{s \in \mathcal{S}} p^s r_{db}^s \right\} = -\max_{d \in \mathcal{D}} \left\{ \frac{CPM_d}{H_d} \bar{r}_{db} \right\}$, for all $b \in \mathcal{B}$. The same solution is obtained for problem ($\bar{\text{D}}$). Therefore, the optimal solution of problem ($\bar{\text{P}}$) is also optimal for the original problem (P). $\square$

Lemma 2.3.3. *If there is never a shortfall in any scenario, then the solution found in Lemma 2.3.1 is optimal to problem (P).*

Proof. If there is never a shortfall, $y_d^{s*} = \bar{y}_d^* = 0$ for all $s \in \mathcal{S}, d \in \mathcal{D}$; therefore, the optimal objective function of both (P) and ($\bar{\text{P}}$) are equal to 0. $\square$

Next, consider the case when there is neither zero shortfall or positive shortfall in all scenarios. Assume in each week the scheduling problem is solved for the deals that run for the next K weeks. Let us analyze the generalization of the random ratings such that instead of having a discrete set of scenarios $\mathcal{S}$, and the associated ratings r_{db}^{sk}, for $k = 1, \dots, K$, it is given a stochastic process R_{db}^k that satisfies Assumption 1 below.

Assumption 1. The ratings stochastic process $\{R_{db}^k : 1 \leq k\}$ is a stationary ergodic process; i.e, its mean and other moments do not change over time and its statistical properties can be inferred from a single sufficiently large random sample.

In this case, the shortfall variables are stochastic decision variables denoted as Y_d^k, $k = 1, \dots, K$; $d \in \mathcal{D}$.

Lemma 2.3.4. *The expected value of the weekly average of the optimal solution when solved for K weeks at a time converges to the one week solution found by replacing the random ratings by their means as K becomes large.*

Proof. Every week the problem is solved for the deals that run for the next K weeks. Consider the generalization of the random ratings that instead of having a discrete set of scenarios $\mathcal{S}$, and the associated ratings r^{sk}_{db}, for $k = 1, \dots, K$, it is given a stochastic process R^k_{db} under general distribution, and the stochastic variable Y^k_d exits. The associated (nonlinear) stochastic programming problem is

$$\begin{array}{llll} \text{P'}(K) & \text{minimize} & \mathbb{E}\dfrac{1}{K}\displaystyle\sum_{k=1}^{K}\sum_{d\in\mathcal{D}} CPM_d Y^k_d & \\ & \text{subject to} & \displaystyle\sum_{d\in\mathcal{D}} H_d x^k_{db} \le L_b & \forall\, b \in \mathcal{B}, k = 1 \dots K, \\ & & \displaystyle\sum_{k=1}^{K}\left(\sum_{b\in\mathcal{B}} R^k_{db} x^k_{db} + Y^k_d\right) \ge K I_d & \forall\, d \in \mathcal{D}, \\ & & x^k_{db} \ge 0 & \forall\, d \in \mathcal{D}, b \in \mathcal{B}; k = 1, \dots, K \\ & & Y^k_d \ge 0 & \forall\, d \in \mathcal{D}; k = 1, \dots, K, \end{array}$$

where $\mathbb{E}$ is the expectation operator. Let's restrict the problem to use the same variables for every week, that is $x^k_{db} = x_{db}$, and $Y^k_d = Y_d$ for all weeks $k = 1, \dots, K$:

$$\begin{array}{llll} \text{P}^{\text{const}}(K) & \text{minimize} & \mathbb{E}\dfrac{1}{K}\displaystyle\sum_{k=1}^{K}\sum_{d\in\mathcal{D}} CPM_d Y_d & \\ & \text{subject to} & \displaystyle\sum_{d\in\mathcal{D}} H_d x_{db} \le L_b & \forall\, b \in \mathcal{B}, \\ & & \displaystyle\sum_{k=1}^{K}\left(\sum_{b\in\mathcal{B}} R^k_{db} x_{db} + Y_d\right) \ge K I_d & \forall\, d \in \mathcal{D}, \\ & & x_{db} \ge 0 & \forall\, d \in \mathcal{D}, b \in \mathcal{B}, \\ & & Y_d \ge 0 & \forall\, d \in \mathcal{D}, \end{array}$$

which is the same as

$$
\begin{aligned}
\mathrm{P}^{\mathrm{const}}(K) \quad \text{minimize} \quad & \mathbb{E}\sum_{d\in\mathcal{D}} CPM_d Y_d \\
\text{subject to} \quad & \sum_{d\in\mathcal{D}} H_d x_{db} \le L_b && \forall\, b\in\mathcal{B}, && (2.15)\\
& \sum_{b\in\mathcal{B}} \left(\frac{1}{K}\sum_{k=1}^{K} R_{db}^{k}\right) x_{db} + Y_d \ge I_d && \forall\, d\in\mathcal{D}, && (2.16)\\
& x_{db} \ge 0 && \forall\, d\in\mathcal{D}, b\in\mathcal{B}, \\
& Y_d \ge 0 && \forall\, d\in\mathcal{D}.
\end{aligned}
$$

Consider the deterministic LP

$$
\begin{aligned}
\mathrm{P}^{\mathrm{det}}(K) \quad \text{minimize} \quad & \sum_{d\in\mathcal{D}} CPM_d y_d \\
\text{subject to} \quad & \sum_{d\in\mathcal{D}} H_d x_{db} \le L_b && \forall\, b\in\mathcal{B}, \\
& \sum_{b\in\mathcal{B}} \bar{r}_{db} x_{db} + y_d \ge I_d && d\in\mathcal{D}, \\
& x_{db} \ge 0 && \forall\, d\in\mathcal{D}, b\in\mathcal{B}, \\
& y_d \ge 0 && \forall\, d\in\mathcal{D}.
\end{aligned}
$$

Recall that Assumption 1 is:

$$
\frac{1}{K}\sum_{k=1}^{K} R_{db}^{k} + \varepsilon(K) = \bar{r}_{db}
$$

where $\varepsilon(K) \to 0$ as $K \to \infty$.

Let $V^*(K)$, $V^{*\mathrm{const}}(K)$ and $V^{*\mathrm{det}}(K)$ be the optimal solutions of problems $P'(K)$, $P^{\mathrm{const}}(K)$ and $P^{\mathrm{det}}(K)$ respectively. Then

$$
V^{*\mathrm{det}}(K) \le V^*(K) \le V^{*\mathrm{const}}(K) \tag{2.17}
$$

The first inequality comes from Lemma 2.3.1 and Assumption 1. The second inequality comes from the fact that the $\mathrm{P}^{\text{const}}(K)$ is a restricted version of P'(K).

Let $H^{\max} := \max_{d\in\mathcal{D}} H_d$, then for any feasible solution the inequality

$$\sum_{b\in\mathcal{B},d\in\mathcal{D}} x_{db} \leq \sum_{b\in\mathcal{B}} L_b/H^{\max}$$

is a valid inequality. Let $M := \sum_{b\in\mathcal{B}} L_b/H^{\max}$.

Let $x_{db}^{*\det}, y_d^{*\det}$ be the optimal solution to the deterministic problem $\mathrm{P}^{\det}(K)$. Plug in this solution into $\mathrm{P}^{\text{const}}(K)$, then constraints (2.15) and (2.16) are met:

$$\begin{aligned} \sum_{d\in\mathcal{D}} H_d x_{db}^{*\det} &\leq L_b && \forall\, b\in\mathcal{B}, \\ \sum_{b\in\mathcal{B}}\left(\frac{1}{K}\sum_{k=1}^{K} R_{db}^k\right) x_{db}^{*\det} + y_d^{*\det} + M\varepsilon(K) &\geq I_d && \forall d\in\mathcal{D}. \end{aligned}$$

Replacing in (2.17), we obtain

$$V^{*\det}(K) \leq V^*(K) \leq V^{*\det}(K) + \sum_{d\in\mathcal{D}} CPM_d M\varepsilon(K) \tag{2.18}$$

therefore, $V^*(K) \to V^{*\det}(K)$ as $\varepsilon(K) \to 0$. $\square$

Accordingly, using the mean of the ratings in the scheduling problem produces a nearly optimal solution if a sufficient number of weeks are optimized at a time and deals run for many weeks. The proofs of Lemmas 1 to 3 are completed by using the primal-dual technique described in [64] of two-stage stochastic programming problems with discrete distributions. Lemma 4 is somewhat new in using more general assumptions compared to the ones found in the literature as well as in the proof technique. Based on the above lemmas, we use the ratings point forecast estimate.

Table 2.1: Small schedule example.

(a) Ratings forecast per break and demographic

Impressions	P6-10	F21-30	M21-30
Break 1	100	300	200
Break 2	200	100	300
Break 3	300	200	100

(b) Targeted demo and weight per advertiser

Advertiser	Target demo	Weight
A1	P6-10	1
A2	F21-30	2
A3	M21-30	3

(c) Original Schedule

Y	P6-10	F21-30	M21-30
Break 1			A3 (3×200)
Break 2	A1 (1×200)		
Break 3		A2 (2×200)	
Total	1200		

(d) Optimized Schedule

X	P6-10	F21-30	M21-30
Break 1		A2 (2×300)	
Break 2			A3 (3×300)
Break 3	A1 (1×300)		
Total	1800 (Lift $\Delta = 600$)		

Note. A spots schedule corresponds to an allocation of spots (advertisements) of different brands to commercial breaks, which are presented in between television shows. Viewers are segmented in demographics and each advertiser targets a specific demographic. In this small example, there are only three breaks, three demographics and three advertisers. Panel (a) displays the forecasted ratings per each demographic on each break. Panel (b) presents the advertiser, their targeted demographic and the weight or importance of each advertiser for the network. Panel (c) is the original schedule, or log, that the network prepares with a heuristic system and provides for optimization. Panel (d) is the post optimized schedule that maximizes the forecasted targeted audience per advertiser multiplied by the corresponding weight. A1 is moved from break 2 to break 3, A2 is moved from break 3 to break 1, and A3 is moved from break 1 to break 2. Unfortunately, such simple moves are not possible due to the many constraints that should be satisfied. However, the real moves that optimize a schedule exhibit chain-like patterns as described in section 2.7.3.

2.4 Three Stages for Scheduling Spots

As discussed in Section 2.3, we carry out the daily scheduling of spots into breaks, and positions within breaks, in three stages. In this section, we provide definitions of variables and describe the models of each stage. For easy reference, we present summary tables of the sets and parameters used in each model. Some of the notation is used in more than one stage. To fix ideas, we present a self-explanatory example in Table 2.1 that depicts the overall approach.

2.4.1 Stage 1: Estimate Deal Weights

Stage 1 answers the following question:

What is the the relative importance of each deal according to: (a) the remaining target demographic impressions of each deal; and (b) the value of these target demographics?

The first stage estimates this importance that we denote as deal's *weight*. The demographic CPM is the cost of reaching 1,000 viewers belonging to the specified audience demographic, a standard measure used by the television industry. This model is solved at the beginning of every week to define each open deal's relative importance, or *weight*, with respect to other open deals. Basically, the model creates a tentative arrangement of spots into breaks for the next week. The arrangement of spots is tentative because the length of all the video spots are not known, some of the advertiser constraints for the complete week may be not defined, and the timetable of breaks may change as the week proceeds. At the moment of solving Stage 1, the available information consists of the following:

- The set of open deals $\mathcal{D}$.
- The set of breaks Θ for the next week. Each break $b \in \Theta$ has a L_b time length.
- The set $\mathcal{Q}$ of demographics. For each demographic $q \in \mathcal{Q}$, the network provides the value for CPM_q, or the cost per thousand impressions based on historical viewership and marketing considerations.
- The deals' specifications are given by the advertisers. These specifications define the targeted demographic of each deal $d \in \mathcal{D}$, which is identified by the binary parameter G_{dq}, for all $q \in \mathcal{Q}$. This is equal to 1 if deal d targets demo q and 0 if not. To simplify the notation, we define the total target CPM of deal d as $CPM_d := \sum_{q \in \mathcal{Q}} G_{dq} CPM_q$, for all $d \in \mathcal{D}$. Also, let $\Theta_{\mathcal{D}}(d) \subset \Theta$ be the set of breaks in which deal d can be shown, for all $d \in \mathcal{D}$; and, for notational convenience as well, we define the set $\mathcal{D}_{\Theta}(b) \subset \mathcal{D}$ of deals that can be scheduled in break b, for all $b \in \Theta$. Also, each open deal has a *target impressions status* I_d for the next week, which is the difference between the promised number of impressions to be delivered on the requested demographic and the impressions already delivered for the deal from the start day, divided by the number

of weeks until the deal expires. In the industry, the Deal Stewardship system closely monitors delivery and provides targets for the remaining life of the deal. In addition, each deal has a restriction on the maximum number of spots that can be broadcasted, J_d. For the next week, some of the video spots are already recorded, so their lengths are determined; however, others are not yet available. Given this lack of information, in order to create the tentative schedule, we use the historic average length of spots $\bar{H}$ in the model.

- To meet the target impressions status, we use a time series forecast model presented in Section 2.6 that estimates ratings $\bar{r}_{qb}$, for each break $b \in \Theta$ and demographic $q \in \mathcal{Q}$. For the sake of notational simplicity, let $\bar{r}_{db} := \sum_{q \in \mathcal{Q}} G_{dq} \bar{r}_{qb}$, for all $d \in \mathcal{D}$.

Table 2.2 presents a summary of the notation described above. As explained before, the schedule cannot be determined exactly because the description of the spots to be scheduled are often incomplete, as several detailed scheduling constraints are missing for spots that are to be aired in the future. Also, the breaks, programs, and lengths are defined but can change significantly over the next few days. Given the available information at this stage, we formulate a linear programming model to generate a tentative multi-day schedule. A set of weights for the day-to-day scheduling problem of Stage 2 is generated as a by-product. An alternative approach would have been to rely upon decision-makers to prioritize deals in a quantitative manner to enable making trade-offs among spots from different deals in Stage 2.

Weight Problem: The decision variables of the Stage 1 model are the x_{db} number of spots of deal d in break b, for all $d \in \mathcal{D}$, and $b \in \Theta_{\mathcal{D}}(d)$; and the y_d shortfall or number of impressions that cannot be delivered of deal d in the case of a lack of forecasted viewership in the target demographic, for all $d \in \mathcal{D}$. The Stage 1 linear programming model is as follows:

$$\begin{aligned}
\text{minimize} \quad & \sum_{d \in \mathcal{D}} CPM_d y_d && && (2.19)\\
\text{subject to} \quad & \sum_{d \in \mathcal{D}_\Theta(b)} x_{db} \leq \tfrac{L_b}{H} && \forall b \in \Theta, && (2.20)\\
& \sum_{b \in \Theta_\mathcal{D}(d)} x_{db} \leq J_d && \forall d \in \mathcal{D}, && (2.21)\\
& \sum_{b \in \Theta_\mathcal{D}(d)} \bar{r}_{db} x_{db} + y_d \geq I_d && \forall d \in \mathcal{D} && (2.22)\\
& y_d, x_{db} \geq 0 && \forall d \in \mathcal{D}, b \in \Theta(d). && (2.23)
\end{aligned}$$

The objective function (2.19) minimizes the value of under-delivered impressions. Constraint (2.20) ensures that the number of spots that are delivered in each break is less than or equal to the number of spots available. Constraint (2.21) ensures that the planned number of spots are allocated as required by the corresponding advertiser of each deal. Constraint (2.22) links the number of spots delivered, the forecasted ratings in the target demographic, the shortfall of impressions, and the target number of impression for each deal. And constraint (2.23) is the nature of variables.

Let W_d be the shadow price of constraint (2.22), for all $d \in \mathcal{D}$. It can be interpreted as the price the network is willing to pay for decreasing the target I_d of deal d by one impression of the target demographic. This is equivalent to treating W_d as the price of each deal-d-impression. We use this price to determine the penalty in the single-day problem that is solved in Stage 2 that we present below.

2.4.2 Stage 2: Schedule Spots in Breaks

Stage 2 allocates the spots to breaks per day. This model schedules spots into the breaks on the actual day, so that the deals assigned higher weights in Stage 1 are given a larger share of the impressions, both in terms of number of spots and preferred time zones per day. For the actual day to be scheduled, the network allows the model to move only

Table 2.2: Stage 1 notation.

Sets and indices:	
$\mathcal{D}, d$	All open deals.
Θ, b	All breaks of the following week. The schedule of breaks is tentative and can change as the week proceeds.
$\mathcal{Q}, q$	Demographics.
$\mathcal{D}_{\Theta}(b) \subset \mathcal{D}$	Deals that can be scheduled in break b, for all $b \in \Theta$.
$\Theta_{\mathcal{D}}(d) \subset \Theta$	Breaks in which deal d can be shown, for all $d \in \mathcal{D}$.
$q(d) \in \mathcal{Q}$	Target demographic for deal d, for all $d \in \mathcal{D}$.
Parameters:	
CPM_q	CPM of demographic q, for all $q \in \mathcal{Q}$.
$\bar{S}$	Average number of spots available in a break.
$\bar{r}_{qb}$	Ratings forecast of break b and demographic q, for all $b \in \Theta, q \in \mathcal{Q}$.
I_d	Target number of impressions for deal d for the following week, for all $d \in \mathcal{D}$.
J_d	Maximum number of deal d spots planned for delivery in the following week, for all $d \in \mathcal{D}$.
Decision Variables:	
x_{db}	Number of spots of deal d in break b, for all $d \in \mathcal{D}$, and $b \in \Theta_{\mathcal{D}}(d)$.
y_d	Shortfall decision variable for deal d , for all $d \in \mathcal{D}$.

a given number of spots to avoid disturbing the elegance of the schedule. Moreover, we include the scheduling rules within each break in Stage 3 to maintain a manageable Stage 2 computational complexity. We focus on optimizing the scheduling at the break level.

Given the Stage 1 deal weights, Stage 2 assigns the spots to breaks in the actual day so that a series of constraints are met, and the ratings per demographic allocated to spots are maximized according to the deal weights. The formulation is a mixed integer linear programing model.

Break Problem: Let $\mathcal{B}$ be the set of breaks on the actual day to be scheduled; let $\mathcal{U}$ be the set of spots; let $\mathcal{D}$ be the set of deals with spots that can be scheduled in the day; and let $\mathcal{Q}$ be the set of audience demographics. For each deal $d \in \mathcal{D}$, $q(d) \in \mathcal{Q}$ is the targeted demographic and $\mathcal{U}_{\mathcal{D}}(d) \subset \mathcal{U}$ is the set of spots that belong to deal d. From this set, only $\hat{\mathcal{U}}_{\mathcal{D}}(d) \subset \mathcal{U}_{\mathcal{D}}(d)$ are guaranteed to be shown on the actual day due to contract considerations. For each spot $i \in \mathcal{U}$, $d(i)$ is the deal to which spot i belongs. The length of spot i is H_i. Given the deal specifications of $d(i)$, $\mathcal{B}_{\mathcal{U}}(i) \subset \mathcal{B}$ is the set of breaks in which spot i can be scheduled. Reciprocally, $\mathcal{U}_{\mathcal{B}}(b) \subset \mathcal{U}$ is the set of spots that can be scheduled in break b, for each break $b \in \mathcal{B}$. The main decision variable of the model is x_{ib}, a binary indicator of whether spot i is in break b, for each spot $i \in \mathcal{U}$ and break $b \in \mathcal{B}_{\mathcal{U}}(i)$. If a spot i cannot be scheduled in any break because of infeasibility, then it is assigned to the bin. Let y_i be the binary decision variable that indicates bin assignation. Recall that Stage 2 receives an original schedule that can be modified by a given maximum number of moves, namely $M^{\texttt{move}}$. Let $\hat{\mathcal{U}}_{\mathcal{B}}(b) \subset \mathcal{U}_{\mathcal{B}}(b)$ be the set of spots scheduled in break b, for each $b \in \mathcal{B}$, and let $\mathcal{U}_{\texttt{bin}} \subset \mathcal{U}$ be the set of spots that are originally in the bin. Accordingly, let $b(i)$ be the scheduled break of spot i, for each $i \in \mathcal{U} \setminus \mathcal{U}_{\texttt{bin}}$. In this given schedule, all the guaranteed spots are scheduled in a break, that is $\hat{\mathcal{U}}_{\mathcal{D}}(d) \cap \mathcal{U}_{\texttt{bin}} = \emptyset$ for all $d \in \mathcal{D}$. At a demand level, i.e., for each deal, the other three relevant dimensions at this stage are the advertisers, the brands, and the product categories. Let $\mathcal{A}$ be the set of advertisers who sign the deals; and, for each advertiser $a \in \mathcal{A}$, let $\mathcal{U}_{\mathcal{A}}(a) \subset \mathcal{U}$ be the set of spots that belong to advertiser a. Each advertiser can have several brands. Let $\mathcal{K}$ be the set of brands; and, for each brand $k \in \mathcal{K}$, let $\mathcal{U}_{\mathcal{K}}(k) \subset \mathcal{U}$ be the set of brand k spots. Finally, across brands, $\mathcal{P}$ is the set of product categories; and, for each $p \in \mathcal{P}$, $\mathcal{U}_{\mathcal{P}}(p) \subset \mathcal{U}$ is the set of spots that belong to product category p. At a supply level, the day is broken into hours $\mathcal{H} = 1, \ldots, 24$, such that $\mathcal{B}_{\mathcal{H}}(h)$ is the set of breaks to be air in hour h, for all $h \in \mathcal{H}$. Finally, for modeling purposes that will become clear later, let $\mathcal{H}_{\mathcal{D}}(d)$ be the set of hours in which the spots of deal d can be shown for all $d \in \mathcal{D}$. Notice that $\mathcal{H}_{\mathcal{D}}(d)$ can be inferred from the set of spots that belong to deal d and the set of hours in which these spots can be shown. Formally, $\mathcal{H}_{\mathcal{D}}(d) := \bigcup_{i \in \mathcal{U}_{\mathcal{D}}(d), b \in \mathcal{B}_{\mathcal{U}}(i)} \left\{h : \underline{h} \leq F_b \leq \bar{h}\right\}$, where $\underline{h}$ and $\bar{h}$ are the start and end day-minutes of hour h, and F_b is the start day-minute of break b.

Objective function (2.24) maximizes the allocated ratings less four types of penalties. The first term is the sum over deals $d \in \mathcal{D}$, guaranteed spots $i \in \hat{\mathcal{U}}_{\mathcal{D}}(d)$, and breaks $b \in \mathcal{B}_{\mathcal{U}}(i)$ of the allocated ratings in the targeted demographic $q(d)$. This equals the ratings per 30 seconds $\bar{r}_{bd} := \bar{r}_{q(d)b}$ multiplied by the ratio of the spot i length to 30 seconds $\frac{H_i}{30}$, multiplied by the allocation variable x_{ib} and multiplied by the weight of the deal W_d, which is obtained from the weight model presented in Subsection 2.4.1. The second term is the penalty for allocating guaranteed spots to the bin, which is the sum of spots assigned to the bin, y_i, multiplied by the ratings allocated in the original schedule $\frac{H_i}{30}\bar{r}_{b(i)d}$, and weighted by W_d. The third term is the sum of decision variable penalties α_{kb} for placing spots of the same brand k scheduled in consecutive breaks b and $b+1$. These penalties are defined in the block of constraints (2.26). Similarly, the fourth term is the sum of decision variable penalties β_{ab} for placing spots of the same advertiser a scheduled in consecutive breaks b and $b+1$, which are defined in the block of constraints (2.27). The fifth term is the sum of decision variable penalties γ_{duh} for not having a uniform scheduled distribution of deal d spots at hour h, defined in the block of constraints (2.28). The parameters $P_{\mathtt{B}}, P_{\mathtt{A}}$ and $P_{\mathtt{V}}$ are input multipliers that calibrate the importance of the penalties in the objective function, and are chosen based on trial and error. The Stage 2 objective function is as follows:

$$\text{maximize} \sum_{\substack{d\in\mathcal{D}, i\in\hat{\mathcal{U}}_{\mathcal{D}}(d),\\ b\in\mathcal{B}_{\mathcal{U}}(i)}} W_d \tfrac{H_i}{30}\bar{r}_{bd}x_{ib} - \sum_{\substack{d\in\mathcal{D},\\ i\in\hat{\mathcal{U}}_{\mathcal{D}}(d)}} W_d \tfrac{H_i}{30}\bar{r}_{b(i)d}y_i - P_{\mathtt{B}}\sum_{\substack{k\in\mathcal{K},\\ b\in\mathcal{B}}}\alpha_{kb} - P_{\mathtt{A}}\sum_{\substack{a\in\mathcal{A},\\ b\in\mathcal{B}}}\beta_{ab} - P_{\mathtt{V}}\sum_{\substack{d\in\mathcal{D},\\ h\in\mathcal{H}_{\mathcal{D}}(d)}}\gamma_{duh}. \tag{2.24}$$

For later reference in Section 3.6, we label the five terms as metrics M1 to M5. Hence, the Stage 2 objective function is equal to M1 - M2 - M3 - M4 - M5. The constraints are the following:

Network feasibility constraints. This set of constraints impose all the rules that define a minimally feasible schedule. For example, each spot has to be scheduled in exactly one break or be placed in the bin. Constraint (2.25a) imposes that each spot $i \in \mathcal{U}$ is scheduled for exactly one break or goes into the bin. Constraint (2.25b) restricts for each break $b \in \mathcal{B}$ that the sum of the spots' length, H_i, scheduled in the break should be less than or equal to the break length, L_b. Constraint (2.25c) imposes that the number of spots that goes into the bin must be less than or equal to $M^{\texttt{bin}}$, the maximum number of spots that are allowed in the bin. Lastly, constraint (2.25d) establishes that the number of spots not moved from the original schedule must be greater than or equal to the total number of spots minus the maximum number of moves allowed, $M^{\texttt{move}}$. The value $M^{\texttt{move}}$ is initially chosen equal to 10% of spots and increased steadily with growing confidence in the model to 50%. This gradual scaling up mitigates the fear that a machine may be unable to generate an elegant schedule—that is a schedule appreciated by experienced schedulers.

$$y_i + \sum_{b \in \mathcal{B}_{\mathcal{U}}(i)} x_{ib} = 1, \qquad \forall i \in \mathcal{U}, \tag{2.25a}$$

$$\sum_{i \in \mathcal{U}_{\mathcal{B}}(b)} x_{ib} H_i \leq L_b, \qquad \forall b \in \mathcal{B}, \tag{2.25b}$$

$$\sum_{i \in \mathcal{U}} y_i \leq M^{\texttt{bin}}, \tag{2.25c}$$

$$\sum_{b \in \mathcal{B}, i \in \hat{\mathcal{U}}_{\mathcal{B}}(b)} x_{ib} + \sum_{i \in \mathcal{U}_{\texttt{bin}}} y_i \geq |\mathcal{U}| - M^{\texttt{move}}. \tag{2.25d}$$

Brand uniform dispersion constraints. It is desirable that spots of the same brand are separated by at least one break of difference. We impose the following constraint to define penalties measured in terms of ratings times deal weights when this requirement is not satisfied. For each brand $k \in \mathcal{K}$ and break $b \in \mathcal{B}$, (2.26a) imposes the relation between the binary variable z_{kb}—which indicates whether at least one brand k spot is scheduled in break b—and the total number of brand k spots in break b; (2.26b) and (2.26c) define the brand separation penalty α_{kb}, which must be greater than or equal to the sum over the spots that are scheduled in consecutive breaks of the delivered ratings $(\frac{H_i}{30}\bar{r}_{bq(d(i))})$, weighted by the deal weight W_d.

$$\sum_{i \in \mathcal{U}_{\mathcal{B}}(b) \cap \mathcal{U}_{\mathcal{K}}(k)} x_{ib} \leq |\mathcal{U}_{\mathcal{K}}(k) \cap \mathcal{U}_{\mathcal{B}}(b)| z_{kb}, \qquad \forall k \in \mathcal{K}, b \in \mathcal{B}, \tag{2.26a}$$

$$\alpha_{kb} \geq \sum_{i \in \mathcal{U}_{\mathcal{B}}(b) \cap \mathcal{U}_{\mathcal{K}}(k)} W_{d(i)} \tfrac{H_i}{30} \bar{r}_{bq(d(i))} (x_{ib} - 1 + z_{kb+1}), \quad \forall k \in \mathcal{K}, b \in \mathcal{B} \setminus \{|\mathcal{B}|\}. \tag{2.26b}$$

$$\alpha_{kb} \geq \sum_{i \in \mathcal{U}_{\mathcal{B}}(b) \cap \mathcal{U}_{\mathcal{K}}(k)} W_{d(i)} \tfrac{H_i}{30} \bar{r}_{bq(d(i))} (x_{ib} - 1 + z_{kb-1}), \quad \forall k \in \mathcal{K}, b \in \mathcal{B} \setminus \{1\}. \tag{2.26c}$$

Advertiser uniform dispersion constraints. These are similar to brand uniform dispersion constraints, but these apply to advertisers instead of brands.

$$\sum_{i \in \mathcal{U}_{\mathcal{B}}(b) \cap \mathcal{U}_{\mathcal{A}}(a)} x_{ib} \leq |\mathcal{U}_{\mathcal{A}}(a) \cap \mathcal{U}_{\mathcal{B}}(b)| v_{ab}, \qquad \forall a \in \mathcal{A}, b \in \mathcal{B}, \tag{2.27a}$$

$$\beta_{ab} \geq \sum_{i \in \mathcal{U}_{\mathcal{B}}(b) \cap \mathcal{U}_{\mathcal{A}}(a)} W_{d(i)} \tfrac{H_i}{30} \bar{r}_{bq(d(i))} (x_{ib} - 1 + v_{ab+1}), \quad \forall a \in \mathcal{K}, b \in \mathcal{B} \setminus \{|\mathcal{B}|\}. \tag{2.27b}$$

$$\beta_{ab} \geq \sum_{i \in \mathcal{U}_{\mathcal{B}}(b) \cap \mathcal{U}_{\mathcal{A}}(a)} W_{d(i)} \tfrac{H_i}{30} \bar{r}_{bq(d(i))} (x_{ib} - 1 + v_{ab-1}), \quad \forall a \in \mathcal{K}, b \in \mathcal{B} \setminus \{1\}. \tag{2.27c}$$

Time based uniform dispersion constraints. Advertisers want spots to be as uniformly distributed as possible in each hour. To model this, we define penalties that measure the deviation in the number of spots per hour from the average number of spots scheduled in an hour. For each deal $d \in \mathcal{D}$ and hour $h \in \mathcal{H}_{\mathcal{D}}(d)$, constraints (2.28a) and (2.28b) define the difference θ_{dh} over the maximum allowed deviation $M_d^{\texttt{dev}}$ from the average number of spots per hour. (2.28c) defines the incurred penalty (γ_{dh}) because of the extra difference.

$$\theta_{dh} + M_d^{\mathsf{dev}} \geq \sum_{\substack{i \in \mathcal{U}_{\mathcal{D}}(d), \\ b \in \mathcal{B}_{\mathcal{U}}(i) \cap \mathcal{B}_{\mathcal{H}}(h)}} x_{ib} - \frac{\sum\limits_{i \in \mathcal{U}_{\mathcal{D}}(d), b \in \mathcal{B}_{\mathcal{U}}(i)} x_{ib}}{|\mathcal{H}_{\mathcal{D}}(d)|}, \qquad \forall d \in \mathcal{D}, h \in \mathcal{H}_{\mathcal{D}}(d), \tag{2.28a}$$

$$\theta_{dh} + M_d^{\mathsf{dev}} \geq \frac{\sum\limits_{i \in \mathcal{U}_{\mathcal{D}}(d), b \in \mathcal{B}_{\mathcal{U}}(i)} x_{ib}}{|\mathcal{H}_{\mathcal{D}}(d)|} - \sum_{\substack{i \in \mathcal{U}_{\mathcal{D}}(d) \\ b \in \mathcal{B}_{\mathcal{U}}(i) \cap \mathcal{B}_{\mathcal{H}}(h)}} x_{ib}, \qquad \forall d \in \mathcal{D}, h \in \mathcal{H}_{\mathcal{D}}(d), \tag{2.28b}$$

$$\gamma_{dh} \geq \theta_{dh} W_{d(i)} \frac{\sum\limits_{b \in \mathcal{B}_{\mathcal{H}}(h)} \frac{H_i}{30} \bar{r}_{bq(d)}}{|\mathcal{B}_{\mathcal{H}}(h)|}, \qquad \forall d \in \mathcal{D}, h \in \mathcal{H}_{\mathcal{D}}(d). \tag{2.28c}$$

A-position and Z-position constraints. By advertisement requirements, some of the spots must be scheduled at the first (last) positions of breaks, which in industry parlance are called A(Z)-position spots. Let $\mathcal{U}_{\mathtt{A}} \subset \mathcal{U}$ be this set of spots. The positioning inside the break is determined in Stage 3, so Stage 2 must ensure that in each break there must be assigned at most one of the spots of $\mathcal{U}_{\mathtt{A}}$. Constraint (2.29a) ensures that this happens. Similarly, some of the spots must be scheduled at the end of a break, or, the Z-position spots. Let $\mathcal{U}_{\mathtt{Z}} \subset \mathcal{U}$ be this set of spots. Constraint (2.29b) imposes that for each break $b \in \mathcal{B}$ only one of these spots can be assigned. The other specific positions inside the breaks are determined by Stage 3, which is explained in Subsection 2.4.3.

$$\sum_{i \in \mathcal{U}_{\mathtt{A}} \cap \mathcal{U}_{\mathcal{B}}(b)} x_{ib} \leq 1, \qquad \forall b \in \mathcal{B}, \tag{2.29a}$$

$$\sum_{i \in \mathcal{U}_{\mathtt{Z}} \cap \mathcal{U}_{\mathcal{B}}(b)} x_{ib} \leq 1, \qquad \forall b \in \mathcal{B}. \tag{2.29b}$$

The other specific positions inside the breaks are determined by Stage 3, which is explained in Subsection 2.4.3.

Minimum separation constraints. The contracts define that certain pairs of spot must be scheduled in breaks separated by at least a given amount of time. For example, an advertiser may want to have their spots separated by at least 30 minutes, or may want a separation of 45 minutes from other spots of the same product category, or may want a separation of one hour from spots of a rival brand. Let $\mathcal{U}^2_{\texttt{Sep}} \subset \mathcal{U} \times \mathcal{U}$ be the set of pair of spot that must be separated; and, for each $(i,j) \in \mathcal{U}^2_{\texttt{Sep}}$, let D_{ij} be the requested separation time, in minutes. Because at this stage we only schedule at a break level and not at a position inside the break, the constraints must ensure that the start time of the breaks where spots i and j are scheduled must be separated by at least D_{ij} plus the average length of a break in minutes, $\bar{L}$. For that, we define the binary decision variable w_{ij} that is equal to 1 if spot i is scheduled before spot j. Constraints (2.30a) and (2.30b) impose that the difference between the start time of the spot that is scheduled later in the day and the spot that is scheduled earlier must be at least $D_{ij} + \bar{L}$, where F_b is the start day-minute of break b.

$$\sum_{b \in \mathcal{B}_{\mathcal{U}}(j)} F_b x_{jb} - \sum_{b \in \mathcal{B}_{\mathcal{U}}(i)} F_b x_{ib} \geq (D_{ij} + \bar{L}) w_{ij} - (24 \times 60)(1 - w_{ij}), \quad \forall (i,j) \in \mathcal{U}^2_{\texttt{Sep}}, \tag{2.30a}$$

$$\sum_{b \in \mathcal{B}_{\mathcal{U}}(i)} F_b x_{ib} - \sum_{b \in \mathcal{B}_{\mathcal{U}}(j)} F_b x_{jb} \geq (D_{ij} + \bar{L})(1 - w_{ij}) - (24 \times 60) w_{ij}, \quad \forall (i,j) \in \mathcal{U}^2_{\texttt{Sep}}. \tag{2.30b}$$

The set $\mathcal{U}^2_{\texttt{Sep}}$ is defined for three different pairs of spot types: 1) separation of spots from the same advertiser or spots from a specific set of advertisers, 2) separation of spots from the same brand or spots from a specific set of brands, and 3) separation of spots from the same product category or spots from a specific set of product categories. If a pair of spots has multiple separation requests, only the constraint for the maximum separation is created.

Association constraints . These are three constraint types required by the advertisers for specific spot pairs related by their content, so they must the broadcasted in a specific order and positions inside a break. The first type is the so called *sandwich constraints*. These spot pairs must be shown in the same break but should be separated by at least one spot within the break, hence the name sandwich. The second type is the *piggyback constraints*. These spot pairs must be shown in the same break in consecutive positions, hence the name "piggyback". For each advertiser $a \in \mathcal{A}$, let $\mathcal{U}^2_{\mathcal{A}\texttt{Sdw}}(a) \subset \mathcal{U}_{\mathcal{A}}(a) \times \mathcal{U}_{\mathcal{A}}(a)$ be the set of spot pairs that must satisfy sandwich constraints, and let $\mathcal{U}^2_{\mathcal{A}\texttt{Pig}}(a) \subset \mathcal{U}_{\mathcal{A}}(a) \times \mathcal{U}_{\mathcal{A}}(a)$ be the spot pairs set that must satisfy piggyback constraints. Constraint (2.31a) imposes that each sandwich and piggyback spot pair must be in the same break. Additionally, constraint (2.31b) ensures that if a sandwich pair (i, j) is scheduled in break $b \in \mathcal{B}_{\mathcal{U}}(i)$, then at least one other spot that is not type A-position or type Z-position must be assigned to the break, in order to have a spot between i and j for the model of Stage 3. Last, the third constraints type is the *consecutive breaks constraints* (2.31c), for which each pair of spots in $\mathcal{U}^2_{\mathcal{A}\texttt{Con}}(a) \subset \mathcal{U}_{\mathcal{A}}(a) \times \mathcal{U}_{\mathcal{A}}(a)$ must be assigned to consecutive breaks.

$$x_{ib} = x_{jb}, \qquad \forall a \in \mathcal{A}, (i, j) \in \mathcal{U}^2_{\mathcal{A}\texttt{Sdw}}(a) \cup \mathcal{U}^2_{\mathcal{A}\texttt{Pig}}(a), b \in \mathcal{B}_{\mathcal{U}}(i), \tag{2.31a}$$

$$\sum_{i' \in \mathcal{U}_{\mathcal{B}}(b)/\{i,j\} \cup \mathcal{U}_{\mathtt{A}} \cup \mathcal{U}_{\mathtt{Z}}} x_{i'b} \geq x_{ib}, \qquad \forall a \in \mathcal{A}, (i, j) \in \mathcal{U}^2_{\mathcal{A}\texttt{Sdw}}(a), b \in \mathcal{B}_{\mathcal{U}}(i), \tag{2.31b}$$

$$x_{ib} = x_{j(b+1)}, \qquad \forall a \in \mathcal{A}, (i, j) \in \mathcal{U}^2_{\mathcal{A}\texttt{Con}}(a), b \in \mathcal{B} \setminus \{|\mathcal{B}|\}. \tag{2.31c}$$

Product category constraints. In each break, a specific maximum number of spots of the same product category can be shown. Let $\mathcal{P}$ be the set of product categories; and, for each $p \in \mathcal{P}$, let $\mathcal{U}_{\mathcal{P}}(p) \subset \mathcal{U}$ be the set of spots that belong to product category p. Constraint (2.32) imposes that a maximum of $M^{\mathtt{B}}_{pb}$ product category p spots can be assigned to break b.

$$\sum_{i \in \mathcal{U}_{\mathcal{P}}(p) \cap \mathcal{U}_{\mathcal{B}}(b)} x_{ib} \leq M^{\mathtt{B}}_{pb}, \qquad \forall p \in \mathcal{P}, b \in \mathcal{B}, \tag{2.32}$$

Nature of variables.

$$\begin{aligned}
x_{ib} &\in \{0, 1\}, & \forall i \in \mathcal{U}, b \in \mathcal{B}_{\mathcal{U}}(i), \\
y_i &\in \{0, 1\}, & \forall i \in \mathcal{U}, \\
z_{kb} &\in \{0, 1\}, \quad \alpha_{kb} \geq 0, & \forall k \in \mathcal{K}, b \in \mathcal{B}, \\
v_{ab} &\in \{0, 1\}, \quad \beta_{ab} \geq 0, & \forall a \in \mathcal{A}, b \in \mathcal{B}, \\
\theta_{dh}, \gamma_{dh} &\geq 0, & \forall d \in \mathcal{D}, h \in \mathcal{H}_{\mathcal{D}}(h), \\
w_{ij} &\in \{0, 1\}, & \forall (i, j) \in \mathcal{U}^2_{\texttt{Sep}}.
\end{aligned} \tag{2.33}$$

Table 2.3: Stage 2 sets and indices.

Supply sets and indices:	
$\mathcal{B}, b$	Breaks within the actual day to be scheduled.
$\mathcal{H}, h$	Hours. $\mathcal{H} = 1, \ldots, 24$
$\mathcal{B}_{\mathcal{H}}(h) \subset \mathcal{B}$	Breaks in hour h, for all $h \in \mathcal{H}$.
$\mathcal{Q}, q$	Demographics.
Demand sets and indices:	
$\mathcal{A}, a$	Advertisers.
$\mathcal{K}, k$	Brands.
$\mathcal{D}, d$	Open deals that can be scheduled in at least one of the breaks $b \in \mathcal{B}$.
$\mathcal{U}, i, j$	Spots.
$\mathcal{P}, p$	Product categories.
$\mathcal{U}_{\mathcal{A}}(a) \subset \mathcal{U}$	Spots that belong to advertiser a, for all $a \in \mathcal{A}$.
$\mathcal{U}_{\mathcal{K}}(k) \subset \mathcal{U}$	Spots of brand k, for all $k \in \mathcal{K}$.
$\mathcal{U}_{\mathcal{D}}(d) \subset \mathcal{U}$	Spots that belong to deal d, for all $d \in \mathcal{D}$.
$d(i) \in \mathcal{D}$	Deal to which spot i belongs, for all $i \in \mathcal{U}$.
$\hat{\mathcal{U}}_{\mathcal{D}}(d) \subset \mathcal{U}_{\mathcal{D}}(d)$	Spots of deal d that are *guaranteed* (i.e., must be aired and together must achieve the viewership target), for all $d \in \mathcal{D}$.
$\mathcal{U}_{\mathcal{P}}(p) \subset \mathcal{U}$	Spots that belong to product category p, for all $p \in \mathcal{P}$.
Sets that define constraints:	
$\mathcal{U}_{\mathtt{A}} \subset \mathcal{U}$	Spots that must be scheduled at the first position within a break*.
$\mathcal{U}_{\mathtt{Z}} \subset \mathcal{U}$	Spots that must be scheduled at the last position within a break*.
$\mathcal{U}^2_{\mathtt{Sep}} \subset \mathcal{U} \times \mathcal{U}$	Pair of spots that must be separated by time*.
$\mathcal{U}^2_{\mathcal{A}\mathtt{Sdw}}(a) \subset \mathcal{U}_{\mathcal{A}}(a) \times \mathcal{U}_{\mathcal{A}}(a)$	Pair of spots that must satisfy sandwich constraint* for advertiser a, for all $a \in \mathcal{A}$.
$\mathcal{U}^2_{\mathcal{A}\mathtt{Pig}}(a) \subset \mathcal{U}_{\mathcal{A}}(a) \times \mathcal{U}_{\mathcal{A}}(a)$	Pair of spots that must satisfy piggyback constraint* for advertiser a, for all $a \in \mathcal{A}$.
$\mathcal{U}^2_{\mathcal{A}\mathtt{Con}}(a) \subset \mathcal{U}_{\mathcal{A}}(a) \times \mathcal{U}_{\mathcal{A}}(a)$	Pair of spots that must be scheduled in consecutive breaks* for advertiser a, for all $a \in \mathcal{A}$.
Demand-Supply sets and indices:	
$\mathcal{B}_{\mathcal{U}}(i) \subset \mathcal{B}$	Valid breaks for spot i (i.e., breaks within which the spot can be aired), for all $i \in \mathcal{U}$.
$\mathcal{U}_{\mathcal{B}}(b) \subset \mathcal{U}$	Valid spots for break b, for all $b \in \mathcal{B}$.
$\hat{\mathcal{U}}_{\mathcal{B}}(b) \subset \mathcal{U}_{\mathcal{B}}(b)$	Spots assigned to break b in the original schedule, for all $b \in \mathcal{B}$.
$b(i) \in \mathcal{B}$	Original break to which spot i is assigned, for all $i \in \mathcal{U} \setminus \mathcal{U}_{\mathtt{bin}}$.
$\mathcal{U}_{\mathtt{bin}} \subset \mathcal{U}$	Spots that are in the bin in the original schedule.
$\mathcal{H}_{\mathcal{D}}(d) \subset \mathcal{H}$	Set of hours within which deal d can be shown, for all $d \in \mathcal{D}$.
$q(d) \in \mathcal{Q}$	Target demographic for deal d, for all $d \in \mathcal{D}$.

Note. Latin calligraphic uppercase denotes set. * Constraints explained in the formulation. The notation $|\mathcal{C}|$ refers to the cardinality of a set $\mathcal{C}$.

Table 2.3 presents the sets and the indices, while Table 2.4 summarizes the parameters used in Stage 2. Table 2.5 presents the decision variables. Where possible, brief notes are given to make the notation understandable without reference to the model.

Table 2.4: Stage 2 parameters.

$\bar{r}_{bq}$	Ratings forecast in 30 seconds of break b, demographic q, for all $b \in \mathcal{B}, q \in \mathcal{Q}$.
F_b	Start time of break b, for all $b \in \mathcal{B}$.
L_b	Length of break b in seconds, for all $b \in \mathcal{B}$.
$\bar{L}$	Average break length in seconds.
H_i	Length of spot i in seconds, for all $i \in \mathcal{U}$.
W_d	Weight of deal d that is obtained in Stage 1, for all $d \in \mathcal{D}$.
$P_{\mathtt{B}}$	Brand separation penalty factor*.
$P_{\mathtt{A}}$	Advertiser separation penalty factor*.
$P_{\mathtt{V}}$	Vertical uniformity penalty factor*.
D_{ij}	Minimum separation* between spots i and j, for all $(i, j) \in \mathcal{U}^2_{\mathtt{Sep}}$.
$M^{\mathtt{bin}}$	Maximum number of spots allowed in the bin. $M^{\mathtt{bin}} \geq \lvert\mathcal{U}_{\mathtt{bin}}\rvert$.
$M^{\mathtt{move}}$	Maximum number of spots that may be moved. Policy variable to ensure the original log's "beauty" is preserved.
$M^{\mathtt{dev}}_d$	Maximum number of spots of deal d per hour allowed to deviate from the average number of spots per hour, for all $d \in \mathcal{D}$.
$M^{\mathtt{B}}_{pb}$	Maximum number of spots from the same product category p that can be placed within break b, for all $p \in \mathcal{P}$ and $b \in \mathcal{B}$.

Note. Latin uppercase denotes parameter. *The extended definition of these parameters are presented in the formulation.

Table 2.5: Stage 2 decision variables.

Binary decision variables:	
x_{ib}	Equal to 1 if spot i is scheduled in break b, 0 otherwise, for all spots $i \in \mathcal{U}$, and breaks $b \in \mathcal{B}_{\mathcal{U}}(i)$.
y_i	Equal to 1 if spot i is added to the bin, 0 otherwise, for all spots $i \in \mathcal{U}$.
z_{kb}	Equal to 1 if at least one spot of brand k is scheduled in break b, 0 otherwise, for all brands $k \in \mathcal{K}$, and breaks $b \in \mathcal{B}$.
v_{ab}	Equal to 1 if at least one spot of advertiser a is scheduled in break b, 0 otherwise, for all advertisers $a \in \mathcal{A}$, and breaks $b \in \mathcal{B}$.
w_{ij}	Equal to 1 if spot i is scheduled before spot j, 0 otherwise, for all pairs of spots $(i, j) \in \mathcal{U}^2_{\mathtt{Sep}}$.
Non negative decision variables:	
α_{kb}	Penalty for scheduling two spots of brand k in consecutive breaks b and $b + 1$, for all brands $k \in \mathcal{K}$ and breaks $b \in \mathcal{B}$.
β_{ab}	Penalty for scheduling two spots of advertiser a in consecutive breaks b and $b + 1$, for all advertisers $a \in \mathcal{A}$ and breaks $b \in \mathcal{B}$.
γ_{dh}	Penalty for vertical uniformity deviation, of deal d and hour h, for all deals $d \in \mathcal{D}$ and hours $h \in \mathcal{H}_{\mathcal{D}}(d)$.
θ_{dh}	Vertical uniformity deviation of deal d and hour h, for all deals $d \in \mathcal{D}$ and hours $h \in \mathcal{H}_{\mathcal{D}}(d)$.

Note. Latin lowercase denotes binary decision variable. Greek letter denotes non-negative decision variable.

The Stage 2 formulation is flexible in terms of accommodating other considerations that may be required by the advertisers or the network. For example, the user may wish to force the scheduling of certain spots that are not guaranteed but must be shown for business considerations, or the user may wish to impose a given number of spots on the log and add constraints that prevent under-delivered deals from becoming worse than before.

2.4.3 Stage 3: Arrange Spots in Break Positions

Stage 3 schedules the spots to their actual positions within the breaks on the actual day. Any unscheduled spots are stored in a repository referred to as the *bin*. After the spots are assigned to breaks, it is necessary to sort them to satisfy the internal break constraints. We can solve the arrangement problem by break. For each break $b \in \mathcal{B}$, let $\tilde{\mathcal{U}}(b)$ be the set of spots that are assigned to break b in Stage 2. Therefore, the number of positions in break b is $|\tilde{\mathcal{U}}(b)|$. Let $\hat{x}_{il}$ be a binary variable that is equal to 1 if spot i is scheduled in position l, for all $i \in \tilde{\mathcal{U}}(b), l \leq |\tilde{\mathcal{U}}(b)|$. To model the piggyback and sandwich constraints, let $\tilde{\mathcal{U}}^2_{\texttt{Pig}}(b)$ be the set of piggyback spot pairs, and let $\tilde{\mathcal{U}}^2_{\texttt{Sdw}}(b)$ be the set of sandwich spot pairs that are assigned to break b in Stage 2. Let $i_{\texttt{A}}(b)$ and $i_{\texttt{Z}}(b)$ be the A-position and Z-position spots that are assigned to break b in Stage 2, if any. Because of all these constraints, there is a chance that the allocation of all the spots to positions is infeasible. In that case, we assign spots to the bin and maximize the weighted ratings allocated, as in Stage 2. Let $\hat{y}_i$ be the binary variable that indicates whether spot $i \in \tilde{\mathcal{U}}(b)$ goes into the bin. Table 2.6 summarizes the notation described above.

The Stage 3 integer programming model, or the *Position Problem*, is as follows:

$$\text{maximize} \sum_{i \in \tilde{\mathcal{U}}(b), l \leq |\tilde{\mathcal{U}}(b)|} W_{d(i)} \tfrac{H_i}{30} \bar{r}_{lq(d(i))} \hat{x}_{il} - \sum_{i \in \tilde{\mathcal{U}}(b)} W_{d(i)} \tfrac{H_i}{30} \bar{r}_{b(i)q(d(i))} \hat{y}_i \tag{2.34}$$

$$
\text{subject to} \quad \sum_{i \in \tilde{\mathfrak{U}}(b)} \hat{x}_{il} \leq 1 \qquad \forall l \leq |\tilde{\mathfrak{U}}(b)|, \tag{2.35}
$$

$$
\sum_{l \leq |\tilde{\mathfrak{U}}(b)|} \hat{x}_{il} + \hat{y}_i = 1 \qquad \forall i \in \tilde{\mathfrak{U}}(b), \tag{2.36}
$$

$$
\hat{x}_{i_{\mathtt{A}}(b)1} = 1, \qquad \hat{x}_{i_{\mathtt{Z}}(b)|\tilde{\mathfrak{U}}(b)|} = 1, \tag{2.37}
$$

$$
\hat{x}_{il} = \hat{x}_{j(l+1)} \qquad l \leq |\tilde{\mathfrak{U}}(b)| - 1, \forall (i,j) \in \tilde{\mathfrak{U}}^2_{\mathtt{Pig}}(b), \tag{2.38}
$$

$$
\hat{x}_{jl} = 0 \qquad l \in \{1,2\}, \forall (i,j) \in \tilde{\mathfrak{U}}^2_{\mathtt{Sdw}}(b), \tag{2.39}
$$

$$
\hat{x}_{il} + \hat{x}_{j(l+1)} \leq 1 \qquad l \leq |\tilde{\mathfrak{U}}(b)| - 1, \forall (i,j) \in \tilde{\mathfrak{U}}^2_{\mathtt{Sdw}}(b), \tag{2.40}
$$

$$
\sum_{t=1}^{l} \hat{x}_{it} \geq \hat{x}_{j(l+2)} \qquad l \leq |\tilde{\mathfrak{U}}(b)| - 2, \forall (i,j) \in \tilde{\mathfrak{U}}^2_{\mathtt{Sdw}}(b), \tag{2.41}
$$

$$
\hat{x}_{il}, \hat{y}_i \in \{0,1\} \qquad i \in \tilde{\mathfrak{U}}(b), l \leq |\tilde{\mathfrak{U}}(b)|. \tag{2.42}
$$

The objective function (2.34) maximizes the weighted ratings allocated to spots as in Stage 2. Constraint (2.35) ensures that no more than one spot is assigned per position, and constraint (2.36) imposes that each spot must be assigned to only one position or to the bin. Constraint (2.37) imposes A-position and Z-position constraints. This means that if Stage 2 assigns the A-position spot $i_{\mathtt{A}(b)}$ to break b, then that spot must be scheduled in the first position of the break. Similarly, if the Z-position spot $i_{\mathtt{Z}(b)}$ is assigned to break b by Stage 2, then that spot must be scheduled in the last position of the break. Constraint (2.38) imposes the piggyback condition, meaning that the spots $(i,j) \in \tilde{\mathfrak{U}}^2_{\mathtt{Pig}}(b)$ must be scheduled in consecutive positions. Constraints block (2.39) to (2.41) are the sandwich constraints: For each pair of sandwich spots $(i,j) \in \tilde{\mathfrak{U}}^2_{\mathtt{Sdw}}(b)$, (2.39) imposes that j can neither be in position 1 nor position 2, constraint (2.40) ensures that spots i and j are separated by at least one position, and constraint (2.41) establishes that spot i must be scheduled before spot j. Finally, constraint (2.42) is the nature of variables.

Table 2.6: Stage 3 sets and indices.

$\mathcal{B}, b$	Breaks.
$\mathcal{U}, i, j$	Spots.
$\tilde{\mathcal{U}}(b) \subset \mathcal{U}$	Spots that are assigned to break b in Stage 2, for all $b \in \mathcal{B}$.
$i_{\mathtt{A}}(b)$	A-position spots that are assigned to break b in Stage 2, if any, for all $b \in \mathcal{B}$.
$i_{\mathtt{Z}}(b)$	Z-position spots that are assigned to break b in Stage 2, if any, for all $b \in \mathcal{B}$.
$\tilde{\mathcal{U}}^2_{\mathtt{Sdw}}(b) \subset \tilde{\mathcal{U}}(b) \times \tilde{\mathcal{U}}(b)$	Pairs of spots that must satisfy sandwich constraints and are assigned to break b in Stage 2, for all $b \in \mathcal{B}$.
$\tilde{\mathcal{U}}^2_{\mathtt{Pig}}(b) \subset \tilde{\mathcal{U}}(b) \times \tilde{\mathcal{U}}(b)$	Pairs of spots that must satisfy piggyback constraints and are assigned to break b in Stage 2, for all $b \in \mathcal{B}$.

2.5 Computational Implementation

The three stages are solved sequentially and with each stage we use a more granular formulation, as discussed in the prior sections. Stage 1 is solved once every week. This stage is solved offline so as to not impact any business operations. However, Stages 2 and 3 are solved one or more times each day. The maximum acceptable solution time is determined by the following two factors: First, all business operations related to the spots for the day in question must be suspended while optimization occurs, and cannot resume until the optimization model has been solved and the revised schedule has been created. Second, the network prefers to perform the optimization after all the work and changes related to the schedule of programs and the spots to schedule have been made, because any changes after Stages 2 and 3 can disturb the optimal solution. In this way, the network can implement the schedule immediately after the optimization. Therefore, there is a limited time for the optimization models to run. Typically, the network allows no more than 10 minutes for solving Stages 2 and 3 combined in order to have a final spot schedule. We describe the solution approach for each stage below.

Stage 1: Estimate Deal Weights. This model is an LP that is easy to solve using a state-of-the-art linear programming solver.

Stage 2: Schedule Spots in Breaks. Initially, we modeled and experimented solving with Stages 2 and 3 in a combined formulation. That formulation did not produce a feasible solution after several hours of branch and bound using a state-of-the-art integer programming solver. So we divided that formulation into Stages 2 and 3. In terms of problem size and computational effort, Stage 2 is the most challenging. For a single day, the typical size of the IP model described in Subsection 2.4.2 includes tens of thousands of variables and constraints (see Table 2.8). The integer programming solver that we have used for this project, Fico Xpress 8.2, produces a near optimal solution after more than three hours, by branch and bound and default settings. Given the limited time for generating a solution and the large problem size, the availability of a starting solution given by the network is an advantage. After experimentation, we found that the solution approach that produces the best solutions quickly is an iterative process in which we feed an initial solution, or warm start solution, and allow a small number of changes. Recall that the network allows a maximum number of moves $M^{\texttt{move}}$, used in constraint (2.25d). As we described before, that number is approximately 50% of the total number of spots. Initially, we set $M^{\texttt{move}}$ equal to 5%, feed the initial solution as warm start producing a new near optimal solution in a few minutes. We then increase $M^{\texttt{move}}$ to 10% using the previously obtained schedule as a warm start to produce a new near optimal solution. The iterative procedure continues in this fashion until reaching the original $M^{\texttt{move}}$. This procedure produces a near optimal solution in more than 30 minutes, which is not quick enough. A second enhancement that accelerates the iterative procedure is to randomly fix certain spots in breaks; then unfixing them and fixing others; until the last iteration has all the spots unfixed (this type of local search procedure had been used in the literature with equal success; for example, by [21]).

The constraints that make the problem difficult to solve are the minimum separation constraints (2.30). Without these constraints the problem solves relatively quickly. These constraints enforce an ordering in every pair of breaks that the network imposes. This difficulty motivates the idea of including in the fixing procedure at each iteration a certain number of the w_{ij} variables that model the ordering between spots i and j, to be equal to the sequence of the warm start solution. If w_{ij} is fixed to one at any iteration, then spot i must be scheduled before spot j and constraint (2.30) associated to the tuple (i, j) is not needed. At the next iteration, this variable is unfixed (this is similar to some extent to [28], who use topological sorting for solving heuristically an open-pit mine production scheduling problem).

Using these enhancements, Xpress computes a near optimal solution at the root node of the branch and bound. It produces a final solution with the original $M^{\texttt{move}}$ in less than 10 minutes.

Stage 3: Arrange Spots in Break Positions. The Stage 3 model that positions the spots inside the breaks is an IP that is easy to solve because the problem can be separated by breaks; therefore, it reduces to arranging about 10 spots in an ordering that satisfies all the constraints.

By reformulating the problem in three stages, the approach solves the scheduling problem while satisfying all the constraints with a near optimal solution in less than 10 minutes. Computational evidence is presented in Section 3.6.

2.6 Ratings Forecasts

Recall that the ratings forecasts are used in the weight problem and the break problem, which we present in Subsections 2.4.1 and 2.4.2, respectively. The parameter to estimate is $\bar{r}_{bq}$, the ratings forecast in break b of demographic q, for all $b \in \mathcal{B}, q \in \mathcal{Q}$. We briefly describe the methods followed and the issues encountered below (for the interested reader, several papers are devoted to ratings forecasting; for example, Danaher and Dagger [25], Danaher et al. [26], Webster et al. [81] and citations within).

Time considerations: We receive the realized ratings one week after the day of broadcasting. Therefore, we have the history of ratings up to one week before the date of airing. Also, a full day is broken into 15–minute intervals; hence, there are 96 quarter hours per day. So, we forecast the ratings per quarter hour and then we map to the corresponding break(s) that are broadcasted in each quarter hour.

Demographics: As we describe in Section 2.1, the demographics that the networks sell are defined using two dimensions: gender and age group. The networks consider 2 possibilities for gender, Female (F) and Male (M) (some media are starting to recognize other genders; for example, Facebook). The age dimension is divided into eight groups (1-5, 6-10, 11-15, 16-20, 21-30, 31-40, 41-51, and 51+). Hence, there are 16 demographics in total ($|\Omega| = 16$). This results in a total of 96×16 time series to forecast.

Ratings Patterns: The ratings are highly variable, as the plots in Figure 2.2 show. Plots 2.2(a) are ratings per day of the week in a representative demographic; each day of the week follows its own pattern. Plots 2.2(b) are the ratings per quarter–hour from 07:00 to 23:45 on a representative day. We can see that the ratings per consecutive quarter–hours are highly correlated, as it would be expected in television programming. A typical quarter– hour–demographic time series has an approximate coefficient of variation (standard deviation/sample mean) of 0.5. Also, when a new season starts with different programming, the past ratings for different programs are often poor predictors for the future. Therefore, as a policy, we restart the forecast every season.

Ratings Models: We tested a battery of time series methods (past ratings average, Holt-Winters and ARIMA); we omit the mathematical details of these models because they are standard. For reference, see any good time series book (e.g., Shumway and Stoffer [66]). We evaluated different strategies of data usage: all past ratings or only a recent subset; all the

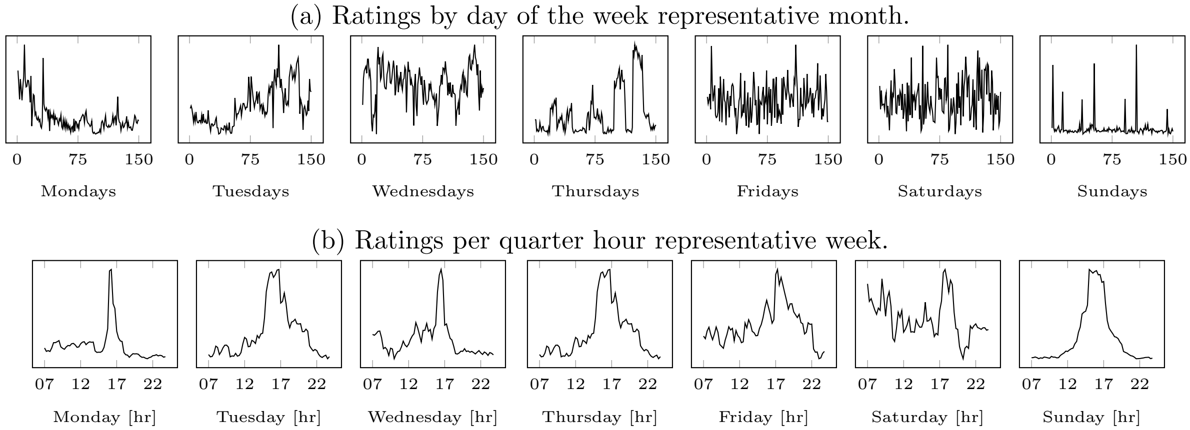


Figure 2.2: Examples of ratings time series.

Note. Ratings are normalized to one and correspond to one particular large demographic.

days of the week combined or only the corresponding day of the week; different aggregation and disaggregation methods (aggregation per day, by show, and by demographic in one or two dimensions). Our benchmark forecast is the past ratings average, which has an approximate 44% mean absolute percentage error (MAPE). Measured by MAPE, none of the methods are found to be superior for all the series. But, for a particular combination of day of the week, time of the day, and demographic, a specific method often produced consistently superior results. We use the method that is the most competitive in terms of MAPE (for test sets). The MAPEs that we obtain for the actual ratings are around 32%. We implemented the time series models using the R forecast package [39].

2.6.1 Protecting Against Uncertainty: Pair of Break Ratings Difference

We further improve the optimization by taking advantage of the correlation between ratings of breaks. The objective function increases by doing swaps that bring the highest increase in delivery impressions among the target demographics. Therefore, more important than the accuracy of point forecast is the precision of the estimate of the difference between break pairs. A simple analysis demonstrates that higher accuracy is obtained by swapping pair of breaks that have higher correlation. Let R_b and $R_{b'}$ be the ratings random variable of breaks b and b'. Assuming that these random variables are normally distributed with mean equal to the point forecasts, $\bar{r}_b$ and $\bar{r}_{b'}$, and standard deviations σ_b and $\sigma_{b'}$, with covariance $\sigma_{bb'}$ and correlation $\rho_{bb'}$, i.e.,

$$\begin{aligned} R_b &\sim \mathcal{N}(\bar{r}_b, \sigma_b^2), \\ R_{b'} &\sim \mathcal{N}(\bar{r}_{b'}, \sigma_{b'}^2), \\ \rho_{bb'} &= \frac{\sigma_{bb'}}{\sigma_b \sigma_{b'}}; \end{aligned}$$

then the difference in ratings random variable, $D_{bb'} \coloneqq R_b - R_{b'}$, is also normally distributed as

$$D_{bb'} \sim \mathcal{N}(\bar{r}_b - \bar{r}_{b'}, \sigma_b^2 + \sigma_{b'}^2 - 2\sigma_{bb'}).$$

Defining $\bar{d}_{bb'} \coloneqq \bar{r}_b - \bar{r}_{b'}$ and $\Delta_{bb'} \coloneqq D_{bb'} - \bar{d}_{bb'}$, we obtain

$$\Delta_{bb'} \sim \mathcal{N}(0, \sigma_b^2 + \sigma_{b'}^2 - 2\sigma_{bb'}). \tag{2.43}$$

From historical data, we observe that the empirical distribution corresponds to the above distribution. For a sample from 24 November 2014 to 27 November 2016, the correlation for every pair of breaks along the day lies in the interval $[-0.061, 0.999]$. High correlation is mostly for pair of breaks that are close to each other, as is shown on the correlation matrix heat map of Figure 2.3(a). Using that historical data, we train the forecast model and produce point forecast, $\bar{r}_b$, for a test week from 12 December 2016 to 18 December 2016 and compute the difference between every pair of breaks $(b.b') : b < b'$ per day, $\bar{d}_{bb'} \coloneqq \bar{r}_b - \bar{r}_{b'}$. After the test week unfolds, we compute the actual difference, $d_{bb'} \coloneqq r_b - r_{b'}$, i.e., the realization of the random variable $D_{bb'}$. Finally, we calculate the difference $\delta_{bb'} \coloneqq d_{bb'} - \bar{d}_{bb'}$, i.e., the realization of random variable $\Delta_{bb'}$. The scatter plot $\delta_{bb'}$ vs. $\rho_{bb'}$ of Figure 2.3(b) shows that the volatility of $\delta_{bb'}$ decreases when $\rho_{bb'}$ increases as distribution (2.43) suggests.

(a) Correlation $\rho_{bb'}$ between pair of breaks (b, b') (b) $\delta_{bb'}$ vs. $\rho_{bb'}$ scatter plot

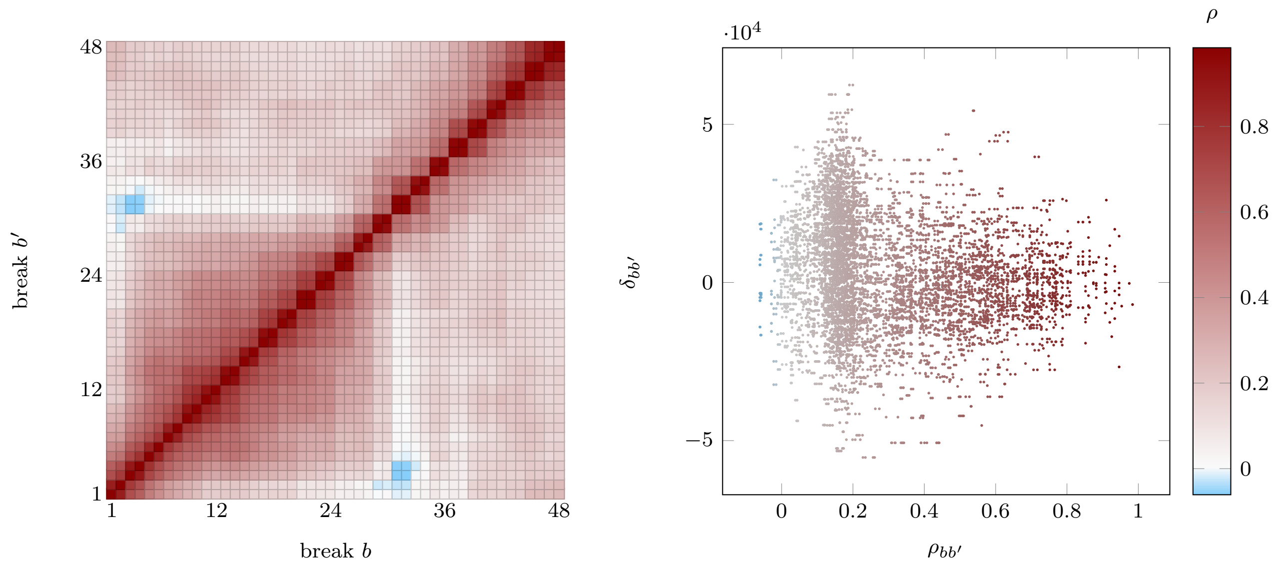


Figure 2.3: Analysis of difference between actual and forecast ratings differences for every pair of breaks.

Note: Panel (a) is a ratings correlation, $\rho_{bb'}$, heatmap between every pair of breaks (b, b'). Breaks are labeled by the half hour of which each break belongs: half hour 1 is 6:00 AM, half hour 2 is 6:30 AM,…, half hour 24 is 5:30 PM,…, and half hour 48 is 5:30 AM. Color blue signals negative correlation and red positive correlation. Higher correlation is mostly between close breaks. In panel (b), $\delta_{bb'} := d_{bb'} - \bar{d}_{bb'}$, where $d_{bb'} := r_b - r_{b'}$ is the difference of actual ratings between pair of breaks (b, b') and $\bar{d}_{bb'} := \bar{r}_b - \bar{r}_{b'}$ is the difference of forecast ratings. Swaps between breaks with low correlation should be avoided because the predictive accuracy of the difference is low.

This result implies that swaps between breaks with high correlation increases the accuracy of the forecasted expected revenue. Managers prefer a schedule that would produce revenue with tight confidence interval rather than a risky schedule with revenue in a wide confidence interval. To exploit the correlation, in Stage 2 we impose constraint (2.44) that prohibits swaps between breaks that have correlation lower than a certain threshold. For every pair of breaks $(b, b') \in \mathcal{B} \times \mathcal{B}$ such that the correlation between them is lower than the minimum threshold ($\rho_{bb'} \leq$ Correlation Threshold), and for every spot i that is assigned to break b in the original schedule, i.e., $i \in \hat{\mathcal{U}}_{\mathcal{B}}(b)$, the spot i cannot be assigned to break b'.

$$x_{ib'} = 0, \qquad \forall b \in \mathcal{B}, i \in \hat{\mathcal{U}}_{\mathcal{B}}(b), b' \text{ such that } \rho_{bb'} \leq \text{ Correlation Threshold.} \tag{2.44}$$

In the actual implementation, the variable $x_{ib'}$ is eliminated from the instance (i.e., $\mathcal{U}_{\mathcal{B}}(b') \leftarrow \mathcal{U}_{\mathcal{B}}(b') \setminus \{i\}$ and $\mathcal{B}_{\mathcal{U}}(i) \leftarrow \mathcal{B}_{\mathcal{U}}(i) \setminus \{b'\}$). This elimination reduces the size of the problem considerably. Simulation experiments show that the expected revenue is not affected greatly by this reduction of flexibility and that the actual revenue is typically close to the predicted revenue. Interestingly, although many possible swaps amongst spots are eliminated, we observe that several long chains of swaps are maintained: starting from any break it is possible to follow a sequence of swaps between connected breaks (pair of breaks that have correlation greater than the threshold) that visits all the breaks and come back to the initial one. The theory of process flexibility and long chains [5, 22, 23, 34, 42, 67, 68, 80] prompts us to think that this phenomena is to be expected, but we keep this question open for further research. We describe the chains in the next section.

2.7 Results

The framework has been used successfully for major television networks of the U.S. and India, generating 3% to 5% increase in revenue which translate in tens of millions of dollars annually for a big network. In order to communicate with the existing networks' data bases and also to integrate the in-house software used for scheduling the logs (before passing a feasible log to the optimizer) with the scheduler, a commercial decision support system was built by a software company [1]. Screen shots of the system are presented in Appendix A.2. As we remark in Section 2.5, we implemented the optimization models in Xpress 8.2.

To quantify the yield generated and the quality of the optimized schedules, in this section we benchmark the schedules produced by our models against the ones provided before optimization for one major network. We compare the daily schedules for a full month. On average, the optimized schedules produce a conservative increase in revenue of more than

[1]The commercial software application is AdVant by RSG Media Systems.

Table 2.7: Stage 2 instance size summary statistics.

	# Breaks $\vert\mathcal{B}\vert$	# Demographics $\vert\mathcal{Q}\vert$	# Advertisers $\vert\mathcal{A}\vert$	# Brands $\vert\mathcal{K}\vert$	# Deals $\vert\mathcal{D}\vert$	# Spots $\vert\mathcal{U}\vert$	# PCat $\vert\mathcal{P}\vert$
Mean	150	16	83	164	124	672	255
Stdev	19.77	2.04	6.50	20.01	18.07	82.62	0.00
CV	0.13	0.12	0.08	0.12	0.15	0.12	0.00

Note. |Set|: number of elements in Set. Stdev: standard deviation. CV: coefficient of variation (Stdev/mean). PCat: product categories. Instances are 31 days of August of a sample year.

$24,500 per day, which translates to almost $9 million per year.

All data, while representative of the problem, are disguised for the purpose of maintaining confidentiality. In this section, we present summary statistics of the instances and the results. The disaggregated data is shown in the online supplement.

2.7.1 Instances, Problem Size, and Optimality Gap

We benchmark the approach using the schedules of one major network's channel for the 31 days of August of a sample year (See Table A.1). In this subsection, we present the mean, standard deviation and coefficient of variation across the 31 days of several measures. The details results are given in Appendix A.1. These results are representative across different cable channels and countries in which the solution is used.

The average day instance has 150 breaks, 16 targeted demographics, 83 advertisers, 164 brands, 124 deals, 672 spots, and 255 product categories. See Table 2.7 for the descriptive statistics of the instances.

Because the Stage 2 model is the most difficult to solve, we describe the problem size of that model. The average problem has 11,661 binary variables, 13,489 continuous variables, and 24,007 constraints. After 10 minutes of optimization, our Stage 2 iterative procedure described in Section 2.5 produces a nearly optimal schedule as indicated by the 0.57% MIP gap on average (100×(LP solution-IP solution)/LP solution). See Table 2.8 for the summary

Table 2.8: Stage 2 problem size summary statistics.

Per day	# Binary vars.	# Continuous vars.	# Constraints	MIP gap
Mean	11,661	13,489	24,007	0.57%
Stdev	7,315	5,704	9,539	0.39%
CV	0.63	0.42	0.40	0.68

Note. Stdev: standard deviation. CV: coefficient of variation (Stdev/mean). MIP gap = ((MIP objective function - best LP bound objective function)/best LP bound objective function)$\times$ 100. Instances are 31 days of August of a sample year.

of the daily problem size (disaggregated metrics are in Table A.2).

2.7.2 Benchmark

In this subsection, we present results showing the day-by-day yield for one month. We set benchmarks using two metrics, namely, the weighted average ratings (M1) and the increase in revenue (Value). The comparison is made between the original schedule provided by the network (ORG) and the schedule obtained after optimization (OPT). The two metrics are:

$$\mathrm{M1} = \sum_{\substack{d\in\mathcal{D}, i\in\hat{\mathcal{U}}_{\mathcal{D}}(d),\\ b\in\mathcal{B}_{\mathcal{U}}(i)}} W_d \frac{H_i}{30} \bar{r}_{bq(d)} x_{ib}; \qquad \mathrm{Value} = \sum_{\substack{d\in\mathcal{D}, i\in\hat{\mathcal{U}}_{\mathcal{D}}(d),\\ b\in\mathcal{B}_{\mathcal{U}}(i)}} CPM_{q(d)} W_d \frac{H_i}{30} \bar{r}_{bq(d)} x_{ib}.$$

We compute the two measures for the ORG and OPT schedules using forecasted as well as actual ratings.

Recall that the weight W_d is an artifact that we compute in order to assign relative importance to the deals as discussed in Section 2.4. However, the network assumes a weight equal to 1 for all deals at the time its original schedule is generated. Hence, in order to establish a fair comparison, in this benchmark we assume $W_d = 1$ for all deals.

Concerning penalties, we compare the sum of the penalties M2+M3+M4+M5 for the ORG and OPT schedules. These penalties are used to impose the soft constraints in the

Table 2.9: Allocated ratings and value based on forecast ratings.

	Weighted average ratings (M1)			Value		
Per day	ORG	OPT	gain	ORG	OPT	gain
Mean	87,193,858	89,088,245	2.21%	\$1,439,032	\$1,468,577	2.08%
Stdev	14,114,964	14,215,505	1.31%	\$279,133	\$283,092	1.30%
CV	0.16	0.16	0.59	0.19	0.19	0.62

Note. Stdev: standard deviation. CV: coefficient of variation (Stdev/mean). ORG: original schedule. OPT: optimized schedule. gain = ((OPT-ORG)/ORG)× 100. We use a random scaling factor to maintain confidentiality, but the order of magnitudes correspond to the true values.

design of the schedule and they do not have an intrinsic monetary value. Therefore, we compare them only for the schedule with forecasted ratings.

Table 2.9 presents the summary results based on the forecasted ratings, which are used to produce the schedules. In terms of allocated ratings, or M1, the optimized schedule allocates, on average, 2.21% higher ratings to the spots. In terms of value, the average day of forecasted ratings has a total value of \$1.47 million, so each percent gain is equivalent to roughly $\$14,700$. For the test month using the forecasted ratings, the OPT schedules produce a 2.08% value gain. Considering the variability of value gain across days using the forecasted ratings, we observe that the coefficient of variation is 0.62, which indicates that the gain is consistently superior (disaggregated results in Table A.3). Evidently, the optimized schedule for the forecasted ratings always dominates the original schedule, which is used as the starting solution.

Table 2.10 displays the mean and standard deviation across the 31 days of the penalties M2+M3+M4+M5 for the ORG and OPT schedules. The penalties of ORG are much higher than OPT, as the 31.49 ratio indicates. Analyzing the disaggregated data per day shown in Table A.4, we observe that the OPT penalties are equal to 0 for many days (15 of 31 days, versus 0 of 31 days for the ORG schedule). Therefore, the OPT schedules violate far fewer of the soft constraints (on some days they violate none of the soft constraints) than the ORG schedules, which always incur penalties due to violations.

Table 2.10: Penalties based on forecast ratings.

	Penalties (M2+M3+M4+M5)		
	ORG	OPT	ORG/OPT
Mean	3,734,878	118,604	31.49
Stdev	2,292,353	181,199	12.65

Note. Stdev: standard deviation. ORG: original schedule. OPT: optimized schedule. We use a random scaling factor to maintain confidentiality, but the order of magnitudes correspond to the true values.

Table 2.11 shows the allocated ratings and value that are obtained by inserting the actual ratings into the schedules produced using the forecasted ratings. In terms of value gain, the OPT schedule has a daily value of \$1.85 million and is, on average, 1.34% superior to ORG, which translates into a daily average of more than \$24,500. However, the value gain coefficient of variation is 1.02, which indicates a high variation. Observing the value gain per each day in Table A.5, the optimized schedule dominates the original schedule in 29 of the 31 days. In the two days on which ORG is better than OPT, the difference is no more than 0.34%. Based on the data, we can compute the empirical probability of a gain using the following expression:

$$\Pr(\text{having a gain}) = \frac{\displaystyle\sum_{\text{day}\in\text{days with gain}} \text{gain}_{\text{day}}\%}{\displaystyle\sum_{\text{day}\in\text{days with gain}} \text{gain}_{\text{day}}\% + \sum_{\text{day}\in\text{days with loss}} \text{loss}_{\text{day}}\%} = 0.9899$$

Therefore, although it is possible that the original schedule is superior compared to the optimized schedule for the actual ratings, the probability of that event is less than 0.0101 for the test month. This unlikely event could occur because the forecasted ratings differ significantly from the actual ratings; and, merely due to chance, the original schedule is better than the optimized schedule. The likelihood of this event shows that the original schedules are of good quality, but because of the highly competitive nature of the U.S. media industry, the networks are constantly searching for sophisticated methods to extract more yield from their audience.

Table 2.11: Allocated ratings and value based on actual ratings.

Per day	Weighted average ratings (M1) ORG	OPT	gain	Value ORG	OPT	gain
Mean	106,047,503	107,570,039	1.50%	$1,825,764	$1,849,022	1.34%
Stdev	22,259,809	22,256,253	1.56%	$500,914	$501,706	1.37%
CV	0.21	0.21	1.03	0.27	0.27	1.02

Note. Stdev: standard deviation. CV: coefficient of variation (Stdev/mean). org: original schedule. opt: optimized schedule. gain = ((opt-org)/org)× 100. I use a random scaling factor to maintain confidentiality, but the order of magnitudes correspond to the true values. Actual ratings are obtained after the spots are aired.

2.7.3 Where does the lift come from?

In this section, we describe the difficulty in obtaining the lift. The obvious question is whether simple heuristics can yield the same quality of solution. For example, if most of the lift comes from exchanging pairs of units, one high viewership unit of an overperforming deal with one low viewership unit of an underperforming deal, then the search for such exchanges becomes relatively trivial. It turns out that the reality of how the lift is created is very different and complicated. We traced the move of a unit from its original position to the next, the unit there to the next, etc., until some unit moves back to the first position thus completing a cycle. To our surprise, the chains of such moves have three features: (i) they are relatively long (see Figure 2.4(c)); (ii) each move is to an adjacent or close break (see Figure 2.4(a)); and (iii) most moves either gain or lose only a little but a few moves create a huge lift indicating that the lift associated with the unit wise moves has a long tail distribution (see Figures 2.4(b) and 2.4(d)). There are two main reasons for this phenomenon: (a) We restrict moves to pairs of break whose ratings are highly correlated; and (b) shorter moves are less likely to affect constraints such as uniformity. The chains structure also allows us to enable manual moves of units when automation of moves is not feasible. As a practical consideration, recall that after the new schedule is determined, it has to be implemented in the system. Software integration is necessary if the software program will make the moves.

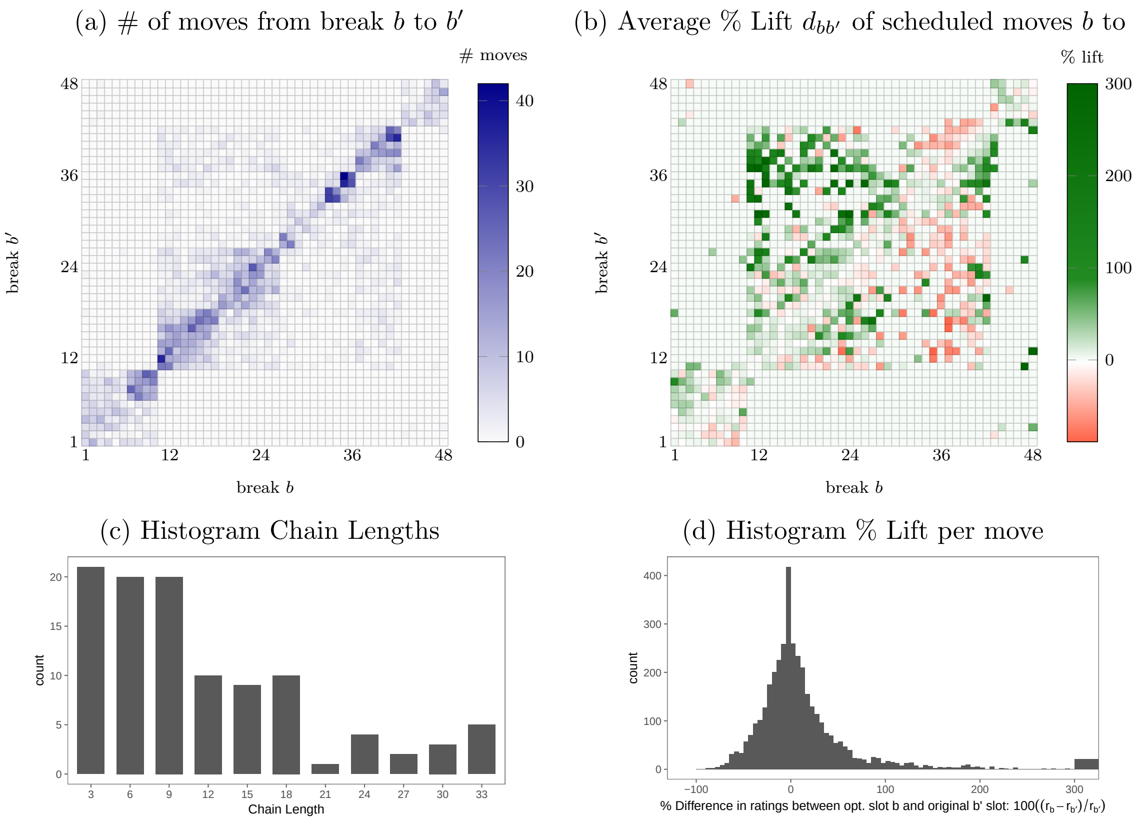


Figure 2.4: Where the lift comes from analysis.

Note. August of a sample year moves. Panel (a) shows the total number of moves from break b to break b'. Breaks are labeled by the half hour of which each break belongs: half hour 1 is 6:00 AM, half hour 2 is 6:30 AM,..., half hour 24 is 5:30 PM,..., and half hour 48 is 5:30 AM. Blue color intensity signals the number of moves, from white marking 0 moves to dark blue signaling up to 42 moves. In total there are 3766 moves. Panel (b) shows the average percentage audience difference (lift) of the scheduled moves from b to b'. Red intensity indicates negative lift, green intensity indicates positive lift and white denotes no lift (it has a right skewed distribution: min. lift is -88.9% , median is 0.0%, mean is 7.8% and max. is 909.9%). We observe that some moves are sacrificed with negative lift from breaks b to b', $b > b'$, with the aim of allowing high lift moves from breaks b to b', $b < b'$. Panel (c) shows the number of chains per chain length. And panel (d) is a histogram of % of lift per move for all the days of August of a sample year. Same distribution is observed if we separate by day of the week or even per each day (see figures A.1 and A.2 in Appendix A.1).

If the moves have to be made manually, then we drop the "last" unit to the bin then work back in the chain moving one unit at a time. Then, if necessary, we bring the dropped (or another) unit back from the bin to the first position.

Table 2.12: Summary statistics of lift per day in %.

Day	count	min	median	mean	max
Monday	445	-80.332	0	8.808	559.6
Tuesday	598	-71.435	0	6.938	239.71
Wednesday	557	-88.902	0.839	20.657	801.081
Thursday	513	-87.135	0.799	17.068	1166.002
Friday	590	-66.559	1.116	13.644	292.541
Saturday	580	-68.083	0	8.849	335.571
Sunday	483	-67.915	3.274	14.471	909.987

2.8 Conclusions

The major revenue source for television networks is the selling of viewers to advertisers. Therefore, an efficient audience distribution among advertisers is essential to maximize the yield. This process involves the scheduling of advertisements, or *spots*, within commercial breaks. The goal of the schedule is to arrange the spots in the breaks so that each spot is shown to its targeted demographic. This multi-period scheduling problem is very difficult to solve, because viewers of different demographics will be watching a particular break at the same time. Thus, at the moment of producing the schedule, various spots are competing for the same break to reach different demographics. On the other hand, the advertisers impose several business restrictions on the schedule, such as minimum separation time between spots of the same product category, and uniform distribution of spots from the same brand.

We designed and implemented a combined solution based on mathematical programming and time series forecasting methods to schedule the spots within breaks in a way that maximizes the value of the audience. The scheduler arranges the spots at the level of positions inside the breaks, which is the maximum level of resolution. The optimization model is a large scale integer programming model. We solve it close to optimality by using an ad-hoc iterative procedure in less than 10 minutes, which is the time available to produce the daily schedule. The schedules are of high quality as measured by standard business metrics and when compared to the mathematical optimal bound. The models are packaged in a full commercial software that is used by leading television networks of the U.S. and India providing increase in revenue of 3% to 5%.

Chapter 3

Peer Effects In The Diffusion Of Solar Panels: A Dynamic Discrete Choice Approach

3.1 Introduction

Installations of solar photovoltaic (PV) systems in the United States have increased rapidly in the last few years, primarily due to federal, state, and local level incentives. In 2006, the federal government established a 30% federal investment tax credit (ITC) that will continue until 2019, when it is scheduled to decline. State support for PV diffusion is also in a downturn. The most ambitious state PV policy, the California Solar Initiative (CSI), is already in its last phase, and it is now available only through municipal service providers. Simultaneously, lower electricity prices in several parts of the U.S., due to low natural gas prices, have further curtailed the penetration of PV.

On the other hand, PV installation costs have dramatically decreased: Average residential prices were less than \$5 per watt in 2013 compared to \$10 per watt in 2007. Figure 3.1 shows the average residential price and existing incentives, in dollars per watt, from 1998 to 2012. We observe that both curves decline with almost the same slope. Therefore, the net installation cost is relatively flat. However, the total annual new solar capacity exponentially increases starting in 2010, as shown in Figure 3.1. How do these price and incentive dynamics impact the diffusion of PV? In particular, among these shifts, how should electric utilities rethink their PV programs when facing budgetary constraints? A better understanding of these issues may be used to design PV programs that targets specific segments of the population.

In this essay, we address these questions by studying the residential PV market in Austin, Texas. We take advantage of a rich data set that includes the demographics, property characteristics, and electricity consumption of each household. We combine these that data with the PV adopter's installation period, installation cost, rebate received, and system attributes. We develop a dynamic discrete choice model (DDCM) that permits the exploration of the effects of various policies and market shifts on the diffusion of PV. Our model characterizes the factors behind PV diffusion at a consumer disaggregated level. It models households as forward-looking consumers. In each period, a household that has not yet adopted PV decides whether to adopt in that period or wait until the next period. In dynamic programming jargon, we formulated the household decision of adoption as an *optimal stopping problem*.

In our formulation, the utility function of each household is a linear combination of the net present value (NPV) of installing a PV system, peer effects, and household heterogeneity. The NPV includes the cost of installation, local rebates, federal ITC, and electricity cost savings. The peer effects constitute the households learning about PV technology from neighbors who have installed PV systems. The household heterogeneity includes the diversity of household's characteristics that are constant in time and of which we do not have data; for example, a predisposition to install renewable energy technology. For estimation of the structural parameters that govern behavior, we design and implement a Bayesian Markov Chain Monte Carlo method that allows us to handle heterogeneity in a dynamic context. We can project the dynamics of the market and conduct a counterfactual analysis because our model is structural, in the sense that its parameters capture the preferences of households and quantifies spatial peer effects. This provides insights about which rebate schedules are more efficient for accelerating the PV diffusion process.

The value of the estimated structural parameters shows that wealthy households are more keen to install a PV system. The parameters also show that peer effects are significant and play an essential role in a household's decision to adopt. Moreover, the marginal effect of a new neighbor adopter is higher for households that are rural, then suburban and finally urban. Furthermore, unobserved household heterogeneity is considerable, as measured by the standard deviation of the random effects that capture heterogeneity.

For validation, we benchmarked our model against the classic Bass model, comparing the predicted number of adopters per quarter in a 2 year test set. Our DDCM had a 6.4% MAPE, which is much lower than the Bass model MAPE of around 80% on the test set. For counterfactual policy analysis, we assumed a hypothetical situation in which the government could adjust the rebate every quarter to match a pre-announced net cost of installation. The net cost is equal to (price charged by the installer − rebate)×(1−ITC).

We created the following net cost scenarios over time: observed net cost as the base case; constant net cost, stepwise decreasing net cost, and stepwise increasing net cost. Each of these net cost-of-installation scenarios translates to rebate schedules that do not deviate significantly from the actual rebate schedule. Subsequently, under each scenario we predict the number of adopters and computed the budget spent on rebates. Scenarios that include a constant or decreasing net cost of installation are Pareto superior to the base case: constant or decreasing scenarios incentivize more adopters than the base case at lower total budget spent in rebates by the government. A constant net cost incentivizes 0.21% more adopters at a 2.65% smaller budget, and a decreasing net cost incentivizes 1.91% more adopters at a 8.46% lower budget. An increasing net cost incentivizes many more adopters than the base case (i.e., 27.63% more adopters), but a much larger government budget spent (i.e., 55.81% larger). A policymaker can use our framework to measure the potential effect of various rebate schedules and adjust it according to her objective.

This essay is structured as follows. Section 3.2 summarizes the related literature. Section 3.3 describes the adoption decision process, the factors that enter into consideration, and data that we collected and leveraged. Section 3.4 specifies the dynamic discrete choice

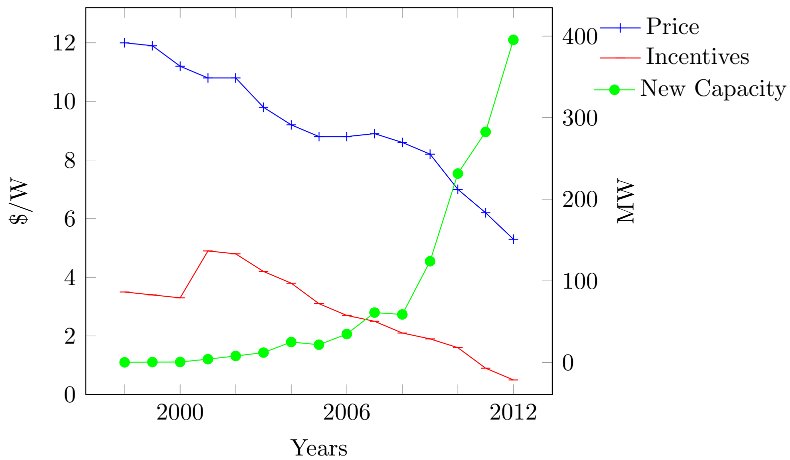


Figure 3.1: Average installed price, incentives and new capacity addition of residential solar panels (capacity lower than 10 kW installed) in the U.S. from 1998 to 2012. 2012 USD.

model. Section 3.5 characterizes the estimation method. Section 3.6 shows the estimation results. Section 3.7 presents the counterfactual analysis, and Section 3.8 concludes the essay.

3.2 Literature Review

The dynamics of the diffusion of new technologies has been studied by researchers from various social disciplines since the mid-20th century. Rogers [60] provides an overview of the applications and methods. The literature on this topic is broad, so we will describe the studies that are most relevant to our research; we will not provide an exhaustive review of the literature on technology diffusion.

The seminal book by Rogers [61] classifies adopters into five categories according to the timing of adoption: (1) *innovators*, (2) *early adopters*, (3) *early majority*, (4) *late majority*, and (5) *laggards*. These categories were defined based on the observation of many empirical examples. Rogers' theory argues that, with the exception of *innovators*, adopters are persuaded by the influence of previous adopters. The theory also argues that the curve of the cumulative number of adopters over time follows an S-shape. This curve initially appears to be linear with a small slope, but the curve then grows exponentially until the number of adopters reaches the diffusion plateau.

Rogers' theory inspires the influential differential equation of the Bass diffusion model (Bass [7]), which formulates the conditional likelihood of adoption over time as the *coefficient of innovation* plus the *coefficient of imitation* multiplied by the function of cumulative adopters. The Bass model has been used successfully to forecast the market penetrations of a variety of products at an aggregate level. Over the years, many refinements have been produced, including decision variables (e.g., optimal pricing policies, advertising), technology innovation, spatial dimension, and network structure (e.g., Bass et al. [8], Van den Bulte and Stremersch [77], Van den Bulte [78], Van den Bulte and Joshi [79], Garber et al. [29], Sood et al. [71, 72], Dover et al. [27], Goldenberg et al. [32, 33], [41], Hu and Van den Bulte [38]). For an in-depth review, see Chandrasekaran and Tellis [20] or Peres et al. [51]. The follow-up approach of Shaikh et al. [65] embeds the Bass model inside a network structure.

Agent-based models (ABMs) constitute a second proposed approach that includes individual decision-making and heterogeneity. These models are microsimulations of the behavior of adopters. Typically, the agents' decision-making rules follow a particular social behavioral theory, and the parameters of the simulation are calibrated from the observed real-world diffusion process (e.g., Kiesling et al. [43], Rahmandad and Sterman [54]). For the PV diffusion context, Palmer et al. [49] use an ABM in the case of Italy. Rai and Robinson [57] study the diffusion of PV in Austin, Texas, by modeling the agents according to the theory of planned behavior Ajzen [3]. We use some elements of this work, which we describe in the following sections.

With a reduced-form specification, Bollinger and Gillingham [15] use data from the California Solar Initiative (CSI) to estimate the probability per ZIP code that a household would adopt a PV system. Their model includes peer effects, and it and provides the theoretical base needed to manage the "reflection problem" of Manski [46]. Essentially, if the data allow us to disentangle the period of the adoption decision from the period of PV installation, then it is possible to infer who influences whom.

Lobel and Perakis [45] use a structural model to analyze the aggregate supply and demand equilibrium of the German PV market. Our research contributes to this stream of work in that we can incorporate in a structural model the peer effects at an individual level. The advantage of using structural modeling is that the formulation reflects the agents' preferences; therefore, the estimated parameters are not dependent on the realization of the variables (e.g., as prices) but only on the decisions made by the agents. This approach allows us to conduct a robust counterfactual analysis.

Our framework follows the structural econometrics methodology initiated by Rust [62]. For a review of this type of work, see, for example, Aguirregabiria and Mira [2]. For estimation, we designed a Bayesian Markov chain Monte Carlo (MCMC) method based on Imai et al. [40] and Norets [48] as described in Section 3.5. Our work is one of the very first that successfully adapts Imai et al. [40] and Norets [48] to a real-world context. From a modeling perspective, our formulation resembles the technology adoption models studied by Ulu and Smith [76] and Smith and Ulu [69, 70]. These authors study the theoretical optimality conditions for the time of adoption under functional assumptions.

3.3 Diffusion and Adoption Decision Process

An interdisciplinary decade-long study of PV diffusion accomplished by the Energy Systems Transformation Research Group of The University of Texas at Austin (see Rai and McAndrews [55], Rai and Robinson [56] and references therein) deduces that a household takes into account the following factors when deciding to adopt.

The first factor is the future cash flow from producing electricity from the PV system. The second factor is a consideration of the dynamic trade-offs between installing the system today (and taking advantage of the available government incentives) versus installing in the future, which may offer fewer incentives. Furthermore, potential adopters consider the dynamic trade-offs between installing the current technology versus waiting for technology improvements to become available, which could offer lower hardware, operation, and maintenance costs. And the third factor is the predisposition to install a PV system. A potential adopter's predisposition changes based on the household's awareness level about the technology and its reliability. We hypothesize that for a particular household, this predisposition increases as the potential adopter sees more households with PV systems. Observing households with PV systems affects potential observers through peer influence (both conscious and unconscious) and because each observed PV system is a signal to the potential adopter that PV technology is reliable. Moreover, the nearer a new installation, the higher the probability that the household will observe the newly installed PV. Therefore, the geographic proximity of new installations plays a significant role in the adoption decision process.

Another interpretation of new installations has a parallel with advertising. Each new installation that the household sees serves as a new advertisement. The more these "ads" the potential adopter views, the more likely she is to decide to adopt a system. For a household that has never seen a solar panel, a PV system could appear unattractive or unusual. However, as the potential adopter sees more systems installed, she becomes more accustomed to their appearance. Eventually, she may develop an interest in this technology.

To model the decision process and empirically estimate the relevance of each factor, we propose a discrete choice dynamic programming model in which the household makes optimal decisions in each period. This approach is superior to a reduced-form analysis in which the adoption is a function of relevant variables, because the estimated parameters do not depend on the actual realization of the data. Therefore, it is possible to conduct a counterfactual analysis of different paths of incentives and technologies.

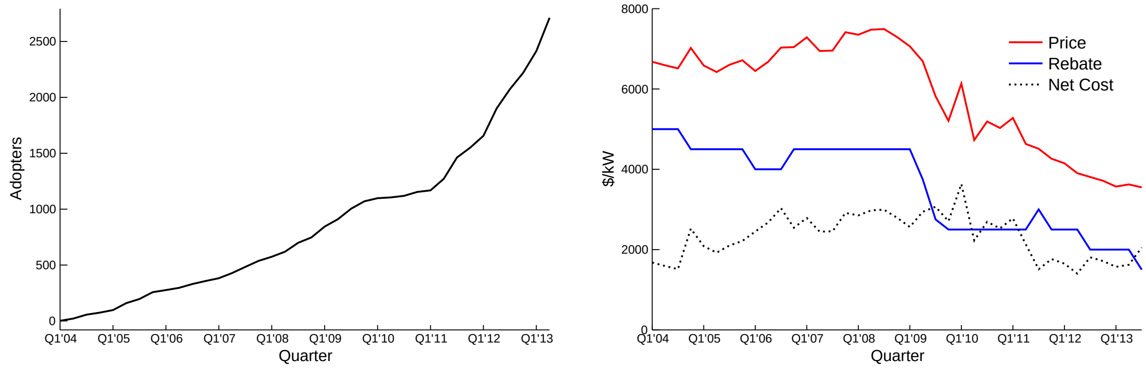
3.4 Dynamic Discrete Choice Model

We index the households by i and denote by $\mathcal{M}$ the set of households that have market potential. We classify the households into two dimensions: the economic set of segments $\mathcal{E}$ and the geographic set of segments $\mathcal{G}$. Let $e_i \in \mathcal{E}$ be the household's economic segment and let $g_i \in \mathcal{G}$ be the household's geographic segment. Time periods of data available are $t \in \{1, \ldots, T\}$ quarters. The households that adopted before or at time t are placed in the set $\mathcal{M}_t \subset \mathcal{M}$, and the time of adoption is t_i for all $i \in \mathcal{M}_T$.

At period t, if household i already adopted, then the household receives bill credits for the electricity generated from its PV system. On the other hand, if the household has not yet adopted, it must decide whether to do so. If the household decides to adopt, then the system is installed, and it starts to produce electricity at time $t+1$ until the end of the system's lifespan (which is assumed to be 20 years, or $L = 80$ quarters).

Given household i's characteristics, the household installs a system of size w_i kW that generates q_{it} (kWh per kW installed) at time t, calculated based on site-specific irradiance. This series is seasonal due to sunlight variability through the year. Furthermore, household i observes the following state variables: the installation price of PV p_t (\$ per kW), the incentive or rebate offered by the utility r_t (\$ per kW), and the residential solar incentive tax credit ITC_t(%) provided by the U.S. Internal Revenue Service[1]. The ITC_t is based on the household's investment—i.e., if i installs a PV system at time t, i would have to pay $(p_t - r_t)(1 - \text{ITC}_t)$ (\$ per kW). Because ITC_t has been constant over time, and because the final net cost of the PV system is the most important factor for the household, we define the per kW net cost of installation, $c_t = (p_t - r_t)(1 - \text{ITC}_t)$, as the data that households observe and consider. Figure 3.2(b) displays the net cost of installation over time. Finally, the household also observes the residential energy solar rate, v_t (\$ per kWh), which corresponds

[1] Energy investment tax credit: http://energy.gov/savings/business-energy-investment-tax-credit-itc.



(a) Accumulated Number of Adopters.

(b) Average Price, Rebate and Net Cost before ITC.

Figure 3.2: Accumulated adopters and cost in Austin, Texas, 2004–2013.

Note. *Price per quarter* is the average installation price that installers charge to households who adopted in each quarter. *Rebate* is the incentive that Austin Energy provided per quarter, and *Net Cost* is Price minus Rebate. Each adopter also obtained the federal incentive tax credit (ITC) of 30% between 2004 and 2013. Source: Austin Energy.

to bill credits for the electricity generated by a PV system.

Additionally, we model the influence of peer effects as follows: At time t, there is an installed base $\mathcal{M}_{t-1} = \{j \in \mathcal{M} : t_j \leq t-1\}$, which consists of all the households that decided to adopt in or before time $t-1$. Notice that the adopters who decide to adopt at time $t-1$ have their PV systems installed at time t. The installed base is known to us, but the household observes some of the installed systems that we assume influence the predisposition of the household to install a PV system. For each household i, we divide $\mathcal{M}_{t-1}$ into subsets of households located a given distance range from i. Let $\mathcal{D} = \{0 < d_1 < d_2 < \ldots < d_D\}$ be the set of discrete distances used for the division (e.g., 0.25 miles, 0.5 miles, 1.0 miles). Let $H_{ilt} \subset \mathcal{M}_{t-1}$ be the set of households located within a distance range $(d_{l-1}, d_l]$ from household i, for all $l = 1, \ldots, D$. Let $h_{ilt} = |H_{ilt}|$ and $\boldsymbol{h}_{it} = \{h_{i1t}, \ldots, h_{iDt}\}$. A similar construction of rings is used by Graziano and Gillingham [35]. For a summary of the notation see Table 3.1.

Table 3.1: Dynamic discrete choice model notation.

Indices and Sets	
$i \in \mathcal{M}$	Set of market potential households.
$e \in \mathcal{E}, g \in \mathcal{G}$	Sets of economic and geographic segments.
$t \in \{1, \ldots, T\}$	Quarters of data available. $T = 39$.
$d_l \in \mathcal{D}$	Set of discrete distances to define neighborhoods by distance radius rings. $\mathcal{D} = \{0 < d_1 < d_2 < \ldots < d_D\}$.
Data	
w_i	PV size (kW) that i would install.
q_{it}	Electricity that would be generated (kWh per kW installed) by household i's PV in period t (calculated based on site electricity irradiance). This series is seasonal due to the sunlight variability throughout the year.
v_t	Residential solar rate ($/kWh). Bill credits for electricity produced by PV at time t.
$e_i \in \mathcal{E}$	Economic segment of household i.
$g_i \in \mathcal{G}$	Geographic segment of household i.
t_i	If i decided to adopt during the period of study, then t_i is the time of adoption. The PV system is installed in the next period. If i did not decide to adopt, then $t_i = \infty$. This applies to all $i \in \mathcal{M}$.
$\mathcal{M}_t \subseteq \mathcal{M}$	Set of households that adopted at or before time t for all $t \in \{1, \ldots, T\}$. $\mathcal{M}_t = \{i \in \mathcal{M} : t_i \leq t\}$, therefore $\mathcal{M}_1 \subseteq \ldots \subseteq \mathcal{M}_T$.
$\mathcal{M}_T^c$	Set of households in the market that have not yet adopted at time T. $\mathcal{M}_T^c = \mathcal{M} \setminus \mathcal{M}_T$.
β	Household discount factor.
L	Lifespan of a PV system (20 years $\times$ 4 quarters per year = 80 quarters).
Decision variable	
$a_{it} \in \{0, 1\}$	Adoption decision at time t of household i. Equals 1 if household adopts and 0 otherwise.
State variables	
Variables observable to both the household and the researcher:	
c_t	PV installation net cost ($/kW) at time t. $c_t = (p_t - r_t)(1 - \text{ITC}_t)$.
H_{ilt}	The set of households that have already installed and are located at a radius distance $d \in (d_{l-1}, d_l]$ from household i. For all $l = 1, \ldots, D$. Also, let the number of adopters $h_{ilt} = \lvert H_{ilt} \rvert$, and $\boldsymbol{h}_{it} = \{h_{i1t}, \ldots, h_{iDt}\}$.
Observable variables to the household, but not to the researcher:	
ξ_i	Additive utility component that reflects individual heterogeneity constant over time. It is a random effect. $\xi_i \sim \mathcal{N}(0, \sigma(\xi))$.
ε_{iat}	Additive utility component for choosing adoption decision a that varies over time.
Structural parameters to estimate	
ρ	Constant term of the utility function.
α_e	Preference of segment e for economic benefit of solar, for all $e \in \mathcal{E}$.
γ_{gl}	Marginal utility of installations that are within radius distance $d \in (d_{l-1}, d_l]$ for all $l = 1 \ldots, D$ and segment $g \in \mathcal{G}$. And let $\boldsymbol{\gamma}_g = \{\gamma_{g1}, \ldots, \gamma_{gD}\}$.
$\sigma(\xi)$	Standard deviation of the random effect ξ.

3.4.1 Per-Period Utility

We formulate the per-period utility using a random utility model. Let a_{it} be the decision variable equal to 1 if household i adopts, or 0 if the household does not adopt. The components that play a role in the utility function are cash flow, peer effects, and heterogeneity between households. We assume that the utility has an additive form composed of the following six components.

The first component is a constant base utility ρ that represents the homogeneous utility of adopting. The second is the net present value from the PV system $\text{NPV}_{it}(c_t)$ multiplied by the preference for economic benefit of solar α_{e_i}, which we consider to be the same for all households in the same economic segment $e_i \in \mathcal{E}$. $\text{NPV}_{it}(c_t)$ equals to the net present value of the bill credits for the electricity generated from the PV system minus the net cost of installation. The third component consists of the peer effects: Each household at a distance between $(d_{l-1}, d_l]$ influences i by $\gamma_{g_i l}$, where g_i is the geographic segment $g_i \in \mathcal{G}$. Let $\boldsymbol{\gamma}_{s_i} = \{\gamma_{g_i 1}, \ldots, \gamma_{g_i D}\}$. Therefore, the total peer effect influence is $\boldsymbol{\gamma}'_{s_i} \boldsymbol{h}_{it}$. The fourth variable ξ_i is the random effect per household, constant over time, that represents the heterogeneity between households not observable to us. This variable can include attitudes toward renewable energy and various idiosyncrasies of the potential adopter. We assume that ξ_i is a normal random variable with mean 0 and standard deviation $\sigma(\xi)$, and we estimate this variable along with the other structural parameters. The sixth component is the variable ε_{i1t}, which is also observable to the household but not to us, which we assume it is an i.i.d. random variable standard extreme value type 1 distributed. Therefore, the utility for adopting is

$$u_{i1t}(c_t, \boldsymbol{h}_{it}, \xi_i, \varepsilon_{i1t}; \rho, \alpha_{e_i}, \boldsymbol{\gamma}_{g_i}) = \rho + \alpha_{e_i} \text{NPV}_{it}(c_t) + \boldsymbol{\gamma}'_{s_i} \boldsymbol{h}_{it} + \xi_i + \varepsilon_{i1t}. \tag{3.1}$$

By assuming that ε_{i1t} is a random variable standard extreme value type I distributed, we impose that its variance is $\pi^2/6$; therefore, all the parameters to estimate $(\rho, \alpha_{e_i}, \boldsymbol{\gamma}_{s_i})$

must be at that scale. If we normalize the utility by any of the parameters (e.g., α_{e_i}) we affect the scale of ε_{i1t}, which would become another parameter to estimate.

We compute the expected profit from the system as follows: Household i pays a net cost c_t, and the household's PV produces electricity from $t+1$ until the end of its lifespan, $t+1+L$. Hence, at every period $t+1 \leq \tau \leq t+1+L$, the household receives $v_\tau q_{i\tau}$ bill credits, which we assume are known to the household. We assume a discount factor of β. Then, the expected profit is

$$\mathrm{NPV}_{it}(c_t) = \left(\sum_{\tau=t+1}^{t+1+L} \beta^{\tau-t} v_\tau q_{i\tau} - c_t \right) w_i. \tag{3.2}$$

If the household decides not to adopt, it obtains a utility component ε_{i0t}, which is observable to the household but not to us, and we assume it is also an i.i.d. random variable standard extreme value type I distributed.

$$u_{i0t}(\varepsilon_{i0t}) = \varepsilon_{i0t} \tag{3.3}$$

3.4.2 Optimal Adoption Decision

Let $\boldsymbol{\varepsilon}_{it} = (\varepsilon_{i0t}, \varepsilon_{i1t})$, and let $\boldsymbol{\theta} = (\rho, \boldsymbol{\alpha}, \boldsymbol{\gamma})$ be the full vector of structural parameters. Then, the decision problem of household i can be formulated as a dynamic programing model. The value function of household i at time t is

$$V_{it}(c_t, \boldsymbol{h}_{it}, \xi_i, \boldsymbol{\varepsilon}_{it}; \boldsymbol{\theta}) = \max_{a \in \{0,1\}} \mathcal{V}_{iat}(c_t, \boldsymbol{h}_{it}, \xi_i, \boldsymbol{\varepsilon}_{it}; \boldsymbol{\theta}), \tag{3.4}$$

where $\mathcal{V}_{iat}$ is the choice specific value function for making the decision $a \in \{0,1\}$:

$$\begin{aligned} \mathcal{V}_{i0t}(c_t, \boldsymbol{h}_{it}, \xi_i, \boldsymbol{\varepsilon}_{it}; \boldsymbol{\theta}) &= u_{i0t}(\varepsilon_{i0t}) + \beta E_{c,\boldsymbol{h},\boldsymbol{\varepsilon}} \left[V_{it+1}(c, \boldsymbol{h}, \xi, \boldsymbol{\varepsilon}; \boldsymbol{\theta}) | c_t, \boldsymbol{h}_{it}, \xi_i, \boldsymbol{\varepsilon}_{it}; \boldsymbol{\theta} \right] \\ &= \varepsilon_{i0t} + \beta E_{c,\boldsymbol{h},\boldsymbol{\varepsilon}} \left[V_{it+1}(c, \boldsymbol{h}, \xi, \boldsymbol{\varepsilon}; \boldsymbol{\theta}) | c_t, \boldsymbol{h}_{it}, \xi_i, \boldsymbol{\varepsilon}_{it}; \boldsymbol{\theta} \right]. \end{aligned} \tag{3.5}$$

$$\begin{aligned} \mathcal{V}_{i1t}(c_t, \boldsymbol{h}_{it}, \xi_i, \boldsymbol{\varepsilon}_{it}; \boldsymbol{\theta}) &= u_{i1t}(c_t, \boldsymbol{h}_{it}, \xi_i, \varepsilon_{i1t}; \boldsymbol{\theta}) \\ &= \rho + \alpha_{e_i} \mathrm{NPV}_{it}(c_t) + \boldsymbol{\gamma}_{g_i}' \boldsymbol{h}_{it} + \xi_i + \varepsilon_{i1t}. \end{aligned} \tag{3.6}$$

Therefore, household i will adopt at t if $\mathcal{V}_{i1t}(c_t, \boldsymbol{h}_{it}, \xi_i, \boldsymbol{\varepsilon}_{it}; \boldsymbol{\theta}) \geq \mathcal{V}_{i0t}(c_t, \boldsymbol{h}_{it}, \xi_i, \boldsymbol{\varepsilon}_{it}; \boldsymbol{\theta})$.

3.4.3 State Variables and Transition Functions

Variables observable at time t:

We model the PV installation net cost ($ per kW) as an AR(1) process:

$$c_{t+1} = \lambda_c^0 + \lambda_c^1 c_t + \epsilon_c, \qquad \epsilon_c \sim N(0, \sigma_{\epsilon_c}^2), \tag{3.7}$$

where λ_c^0, λ_c^1 and σ_{ϵ_c} are computed from historical data. Figure 3.2(b) presents the time series c_t from Q1-2001 to Q4-2013.

Variables observable only to households at time t:

Two variables are observable only to households. The first variable is the heterogeneity among households ξ_i which is constant over time. This variable could include attitudes toward renewable energy or idiosyncrasy. The second variable is the latent utility that is variable among households, time, and decision of adoption, ε_{iat}. We model this as an i.i.d. standard extreme value type I distribution.

3.5 Bayesian Estimation

For estimation, we designed a Bayesian MCMC algorithm based on a combination of the methods developed by Imai et al. [40] and Norets [48]. This algorithm conducts a Metropolis–Hastings with a Gibbs sampler, in which at iteration r it draws a candidate parameter $\boldsymbol{\theta}^{(r)} = \{\rho^{(r)}, \boldsymbol{\alpha}^{(r)}, \boldsymbol{\gamma}^{(r)}\}$; a random grid of size J, $\{\boldsymbol{c}^{j(r)}\}_{j=1}^{J}$; and it approximates the value function using N past sampler iterations. Let $\hat{\mathcal{V}}_{iat}^{(r)}(c_t^{j(r)}, \boldsymbol{h}_{it}, \xi_i^{(r)}; \boldsymbol{\theta}^{(r)})$ be the approximation of the value function by making decision $a \in \{0, 1\}$ at iteration r given state. Let $\hat{V}_{it}^{(r)}(c_t^{j(r)}, \boldsymbol{h}_{it}, \xi_i^{(r)}; \boldsymbol{\theta}^{(r)})$ be the approximation of the value function at iteration r given state; and let $\hat{E}\left[\hat{V}_{it+1}^{(r)}(c, \boldsymbol{h}_{it+1}, \xi; \boldsymbol{\theta}) | c_t^{j(r)}, \boldsymbol{h}_{it}, \xi_i^{(r)}; \boldsymbol{\theta}^{(r)}\right]$ be the approximation of the expectation of the value function in the next period at iteration r. Then the algorithm in Table 3.2

computes the approximations as:

$$\hat{E}\left[\hat{V}_{it+1}^{(r)}(c,\boldsymbol{h}_{it+1},\xi;\boldsymbol{\theta})|c_t^{j(r)},\boldsymbol{h}_{it},\xi_i^{(r)};\boldsymbol{\theta}^{(r)}\right]=\sum_{n=r-N}^{r-1}\sum_{j=1}^{J}\frac{\hat{V}_{it+1}^{(n)}(c_{t+1}^{j(n)},\boldsymbol{h}_{it+1},\xi_i^{(n)};\boldsymbol{\theta}^{(n)})f(c_{t+1}^{j(n)}|c_t^{j(r)})}{\sum_{n=r-N}^{r-1}\sum_{j=1}^{J}f(c_{t+1}^{j(n)}|c_t^{j(r)})}.$$

$$\hat{\mathcal{V}}_{i0t}^{(r)}(c_t^{j(r)},\boldsymbol{h}_{it},\xi_i^{(r)};\boldsymbol{\theta}^{(r)})=\beta\hat{E}\left[\hat{V}_{it+1}^{(r)}(c,\boldsymbol{h}_{it+1},\xi;\boldsymbol{\theta})|c_t^{j(r)},\boldsymbol{h}_{it},\xi_i^{(r)};\boldsymbol{\theta}^{(r)}\right].$$

$$\hat{\mathcal{V}}_{i1t}^{(r)}(c_t^{j(r)},\boldsymbol{h}_{it},\xi_i^{(r)};\boldsymbol{\theta}^{(r)})=\rho^{(r)}+\alpha_{e_i}^{(r)}\mathrm{NPV}_{it}(c_t^{j(r)})+\boldsymbol{\gamma}_{s_i}^{(r)}\boldsymbol{h}_{it}+\xi_i^{(r)}.$$

$$\hat{V}_{it}^{(r)}(c_t^{j(r)},\boldsymbol{h}_{it},\xi_i^{(r)};\boldsymbol{\theta}^{(r)})=\exp\left(\log\hat{\mathcal{V}}_{i0t}^{(r)}(c_t^{j(r)},\boldsymbol{h}_{it},\xi_i^{(r)};\boldsymbol{\theta}^{(r)})+\log\hat{\mathcal{V}}_{i1t}^{(r)}(c_t^{j(r)},\boldsymbol{h}_{it},\xi_i^{(r)};\boldsymbol{\theta}^{(r)})\right).$$

$$\hat{p}_{iat}^{(r)}(\boldsymbol{\theta}^{(r)})=\frac{\exp\hat{\mathcal{V}}_{iat}^{(r)}(c_t^{j(r)},\boldsymbol{h}_{it},\xi_i^{(r)};\boldsymbol{\theta}^{(r)})}{\exp\hat{\mathcal{V}}_{iat}^{(r)}(c_t^{j(r)},\boldsymbol{h}_{it},\xi_i^{(r)};\boldsymbol{\theta}^{(r)})+\exp\hat{\mathcal{V}}_{ia't}^{(r)}(c_t^{j(r)},\boldsymbol{h}_{it},\xi_i^{(r)};\boldsymbol{\theta}^{(r)})}.$$

The full market size is too large to be handled computationally in the likelihood function. We aggregate the households that did not yet adopt into clusters of households that share similar characteristics (specifically, home market value, size of the house, tree-cover, and irradiance received) per block group using the k-means clustering technique in order to reduce the computational complexity of the estimation algorithm.

Figure 3.3 shows the households before and after aggregation. Panels 3.3a and 3.3b show the Austin Energy service area, in which the red dots represent the adopters during the period of analysis. Each black dot represents the following: A potential adopter in Panel 3.3a, and a cluster of potential adopters in Panel 3.3b. Panel 3.3c presents a zoom into a specific block group in which each black dot symbolizes a potential adopter and each green dot represents a cluster of potential adopters. Then, in the likelihood function we weight each cluster by the number of potential adopters that belongs to that cluster.

Table 3.2: Algorithm MCMC estimation.

Input : *Data*: household characteristics, time of adoption, costs, value of solar. β discount factor. R MCMC iterations. shape, base scale, $\sigma_\rho, \sigma_\alpha$ and σ_γ hyperparameters to control acceptance rate and convergence of MCMC.

Output : $\bar{\boldsymbol{\theta}} = (\bar{\rho}, \bar{\alpha}, \bar{\gamma}, \bar{\sigma}_\nu)$ posterior mean of structural parameters. For the sake of exposition, assume only one economic and one geography segment.

Variables : $\rho^{(r)}, \alpha^{(r)}, \gamma^{(r)}, \nu_i^{(r)}$ accepted values at iteration r. $\rho', \alpha', \gamma', \nu_i'$ proposal. $\rho^0, \alpha^0, \gamma^0, \nu_i^0 = 0$

```
for r ∈ 1 to R do
    // ********** HOUSEHOLDS RANDOM EFFECTS **********
    c, f = SimulateCostFunction()
    for i ∈ Households do
        Ê_i = EmaxApprox(V̂_i, f)
        l_i = LogLik(ρ^(r−1), α^(r−1), γ^(r−1), ν_i^(r−1), Ê_i, Data, β, c, f)
        scale = base scale + (∑ ν_i^2) /2
        ξ ∼ InverseGamma(shape, scale)
        σ_ν = √ξ
        ν_i' ∼ Normal(0, σ_ν)
        l_i' = LogLik(ρ^(r−1), α^(r−1), γ^(r−1), ν_i', Ê_i, Data, β, c, f)
        u ∼ U(0, 1)
        if log(u) ≤ min(l_i' − l_i, 0) then ν_i^(r) = ν_i' else ν_i^(r) = ν_i^(r−1)
        V̂_i = VApprox(ρ^(r−1), α^(r−1), γ^(r−1), ν_i^(r), Ê_i, Data, β, c, f)
    // ********** BASE UTILITY **********
    c, f = SimulateCostFunction()
    for i ∈ Households do
        Ê_i = EmaxApprox(V̂_i, f)
        l_i = LogLik(ρ^(r−1), α^(r−1), γ^(r−1), ν_i^(r), Ê_i, Data, β, c, f)
    l = ∑ l_i
    ρ' ∼ Normal(ρ^(r−1), σ_ρ)
    for i ∈ Households do  l_i' = LogLik(ρ', α^(r−1), γ^(r−1), ν_i^(r), Ê_i, Data, β, c, f)
    l' = ∑ l_i'
    u ∼ U(0, 1)
    if log(u) ≤ min(l' − l, 0) then ρ^(r) = ρ' else ρ^(r) = ρ^(r−1)
    for i ∈ Households do  V̂_i = VApprox(ρ^(r), α^(r−1), γ^(r−1), ν_i^(r), Ê_i, Data, β, c, f)
    // ********** ECONOMIC FACTOR **********
    c, f = SimulateCostFunction()
    for i ∈ Households do
        Ê_i = EmaxApprox(V̂_i, f)
        l_i = LogLik(ρ^(r), α^(r−1), γ^(r−1), ν_i^(r), Ê_i, Data, β, c, f)
    l = ∑ l_i
    α' ∼ Normal(α^(r−1), σ_α)
    for i ∈ Households do  l_i' = LogLik(ρ^(r), α', γ^(r−1), ν_i^(r), Ê_i, Data, β, c, f)
    l' = ∑ l_i'
    u ∼ U(0, 1)
    if log(u) ≤ min(l' − l, 0) then α^(r) = α' else α^(r) = α^(r−1)
    for i ∈ Households do  V̂_i = VApprox(ρ^(r), α^(r), γ^(r−1), ν_i^(r), Ê_i, Data, β, c, f)
    // ********** PEER EFFECTS **********
    c, f = SimulateCostFunction()
    for i ∈ Households do
        Ê_i = EmaxApprox(V̂_i, f)
        l_i = LogLik(ρ^(r), α^(r), γ^(r−1), ν_i^(r), Ê_i, Data, β, c, f)
    l = ∑ l_i
    γ' ∼ Normal(γ^(r−1), σ_γ)
    for i ∈ Households do  l_i' = LogLik(ρ^(r), α^(r), γ', ν_i^(r), Ê_i, Data, β, c, f)
    l' = ∑ l_i'
    u ∼ U(0, 1)
    if log(u) ≤ min(l' − l, 0) then γ^(r) = γ' else γ^(r) = γ^(r−1)
    for i ∈ Households do V̂_i = VApprox(ρ^(r), α^(r), γ^(r), ν_i^(r), Ê_i, Data, β, c, f)

ρ̄ = mean(ρ^(r)); ᾱ = mean(α^(r)); γ̄ = mean(γ^(r)); σ̄_ν = mean(σ_ν^(r));
```

(a) No aggregation. (b) Aggregation. (c) Example of households block group aggregation.

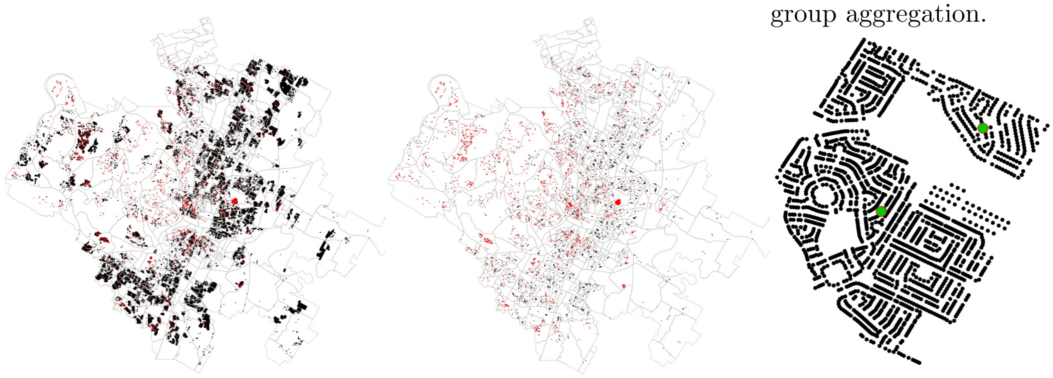


Figure 3.3: Solar adopters and non-adopters until 2013.

Source: Austin Energy. Black dots represents non-adopters and red dots are adopters in the period of analysis. The use of clusters on the likelihood function is an idea inspired by the weighted exogenous sampling maximum likelihood method of Manski and Lerman [47].

3.6 Estimation Results

The utility function presented in Equation (3.1) has four dimensions that can be adjusted to test robustness, improve forecast accuracy, and refine policy recommendations via counterfactual analysis: (a) the inclusion or exclusion of peer effects, (b) the inclusion or exclusion of random effects, (c) economic segment cohorts, and (d) radial distance brackets (if peer effects are admitted). We assume the set of geographic segments defined by the Claritas PRIZM segmentation: urban, suburban, and rural segments. For each model specification, we estimate the corresponding parameters using the Bayesian method explained in Section 3.5. We compare the models using two main metrics: in-sample deviance information criterion (DIC)[2], and out-of-sample mean absolute percentage error (MAPE). The dimensions values that we consider are:

(a) **Peer effects.** With or without peer effects.

[2] DIC $= -2\log \boldsymbol{\ell}(\bar{\boldsymbol{\theta}}) + 2p$, where, $\bar{(.)}$ is the mean operator, and p is the effective number of parameters. This can be computed as $p_D = 2(\log \boldsymbol{\ell}(\bar{\boldsymbol{\theta}}) - \overline{\log \boldsymbol{\ell}(\boldsymbol{\theta})})$ or $p_V = 2\text{var}\,(\log \boldsymbol{\ell}(\boldsymbol{\theta}))$. [31, 73]

(b) **Random effects.** Nonrandom effects (NRE) or random effects (RE).

(c) **Economic segment cohorts.** We proxy the cohorts by the market value of the houses. We select four different sets of cutoff values (in thousands of dollars) to define the segments: (0,
Inf); (0, 300, *Inf*); (0, 300, 600, *Inf*); and (0, 150, 300, 500, *Inf*). *Inf* has a very large value.

(d) **Radial distance brackets.** The cutoffs (in miles) that we consider are: (0, 0.5); (0, 0.5, 1); (0, 0.5, *Inf*); (0, 1, *Inf*); and (0, *Inf*). *Inf* correspond to the limits that Austin Energy serves and where the PV rebates are available. This dimension applies only for models with peer effects.

We divide the data in *training set* and *test set* as follows. Let $\hat{T} < T$, a time period before the last observed period. The set of adopters until time $\hat{T}$ is $\mathcal{M}_{\hat{T}} = \{i \in \mathcal{M} : t_i \leq \hat{T}\}$. We fit the model using the adopters in $\mathcal{M}_{\hat{T}}$, then compute the out-of-sample predictive fit for the households in $\mathcal{M}^c_{\hat{T}} = \mathcal{M} \setminus \mathcal{M}_{\hat{T}}$. Let S simulations draws $\{\boldsymbol{\theta}^s_{\hat{T}}\}_{s=1}^S$ of the parameters, fitting the model to the adopters in $\mathcal{M}_{\hat{T}}$. Let N_t be the number of adopters at time t. Let $\mathcal{M}^c_t$ be the set of households who have not adopted until time t. Then, the foretasted number of adopters at time t, $\hat{N}_t$, is

$$\hat{N}_t = \frac{1}{S}\sum_{s=1}^{S}\sum_{i\in\mathcal{M}^c_t}\left(\hat{p}^s_{i1t}(\boldsymbol{\theta}^s)\prod_{\tau=1}^{t-1}\hat{p}^s_{i0\tau}(\boldsymbol{\theta}^s)\right).$$

Let $\mathcal{Q} := \{\text{2011-Q1}, \ldots, \text{2013-Q2}\}$ be the set of quarters to predict.

Table 3.3 presents a summary of the various model specifications and their error metrics. DIC is in the range of 24,700 to 27,893, without distinguishing a clear best model. On the other hand, MAPE fluctuates between 6.40% and 42.84%, which indicates that model VI-RE is the best, with minimum MAPE. Before discussing that model in detail, an examination of these dimensions reveals some important insights. In the following paragraphs, all percentage numbers refers to MAPE unless it is stated otherwise.

First, the models without peer effects are the worst. That is, specification XII-NRE has 42.84% and specification XII-RE has 41.88%, versus the closest model with peer effects IX-NRE, which has 36.74%. Therefore, peer effects are an important factor in the decision to adopt.

Second, among the models that include peer effects, the models RE dominates models NRE. The exception is Model III, in which III-RE 10.20% is worse than III-NRE 9.15%.

Third, regarding the economic segment cohorts, the optimal segmentation is to divide the households into two cohorts around the \$300,000 cutoff value. One cohort (0, *Inf*) is dominated in every comparison versus two cohorts. For example, when we control for radial distance brackets at (0, 0.5) and RE, model I-RE has 11.03% versus model IV-RE , which has 10.95%. On the other hand, more than two cohorts is worse than only two cohorts.

Fourth, radial distance brackets (0, 1, *Inf*) dominates all other partitions of the distance space, when controlling for home market value cohorts and the inclusion or exclusion of random effects.

This analysis concludes that the best model is composed of (a) two cohorts of market value divided at the \$300,000 line, (b) radial distance brackets up to one mile and more than one mile, and (c) the inclusion of random effects. Model VI-RE meets these criteria, and it has a MAPE of 6.40%. Table 3.4 presents the estimated parameters of this model, which are equal to the posterior mean of the MCMC distribution, the Bayesian 95% credible intervals of the parameters, Geweke p-value to test convergence, and the acceptance rate.

The base utility ρ is a reference at which all the households initially value PV. On model VI-RE this base utility equals -6.83860. That is the log odds of a *hypothetical* household with NPV equal to zero, no adopter neighbors, and heterogeneity also equal to

Table 3.3: Summary of models and metrics.

	α	γ	DIC		MAPE (%)	
Model	Home Mkt Value Cohorts ($1K)	Radial Distance Brackets (Miles)	NRE	RE	NRE	RE
I	(0, Inf)	(0, 0.5)	27,893	27,020	13.93	11.03
II	(0, Inf)	(0, 0.5, 1)	26,441	26,330	25.24	21.76
III	(0, Inf)	(0, 1, Inf)	26,091	25,907	9.15	10.20
IV	(0, 300, Inf)	(0, 0.5)	26,443	26,698	10.95	7.94
V	(0, 300, Inf)	(0, 0.5, 1)	26,921	26,793	21.56	18.46
VI	(0, 300, Inf)	(0, 1, Inf)	26,242	26,260	7.56	6.40
VII	(0, 300, Inf)	(0, 0.5, Inf)	26,593	26,646	12.22	11.80
VIII	(0, 300, 600, Inf)	(0, 0.5)	27,044	27,055	24.89	22.16
IX	(0, 300, 600, Inf)	(0, 0.5, 1)	26,681	26,524	36.74	32.58
X	(0, 150, 300, 500, Inf)	(0, Inf)	24,838	24,700	21.27	20.60
XI	(0, 150, 300, 500, Inf)	(0, 1, Inf)	26,327	25,750	15.29	14.79
XII	(0, 150, 300, 500, Inf)	No Peer Effects	24,875	24,760	42.84	41.88

Note. NRE: model without random effects. RE: model with random effects.

0. The odds of such a hypothetical household adopting PV is $\exp(-6.83860) = 0.00107160$, and the probability of adopting is $\exp(-6.83860)/(1 + \exp(-6.83860)) = 0.00107046$. The households whose houses have a market price that exceeds $300,000 finds more value in a PV system than households with houses that have a lower market price. The parameter $\alpha_{[300\text{K},Inf)}$ is 0.00056 versus $\alpha_{[0,300\text{K})}$ at 0.00028 level. This difference reflects that households with houses of higher value want to invest more in their houses than the owners of less expensive properties.

Next, for all geographic segments, the peer effect γ of the closest installations (i.e., less than a mile away) is much higher than the effect of further installations. For urban locations, $\gamma_{.,\text{urban}}$ is 0.03374 versus -0.00026. For suburban locations, $\gamma_{.,\text{suburban}}$ is 0.04702 versus 0.00005. For rural locations, $\gamma_{.,\text{rural}}$ is 0.13716 versus -0.00033. This negative values on urban and rural households more than a mile away are very close to zero and have a

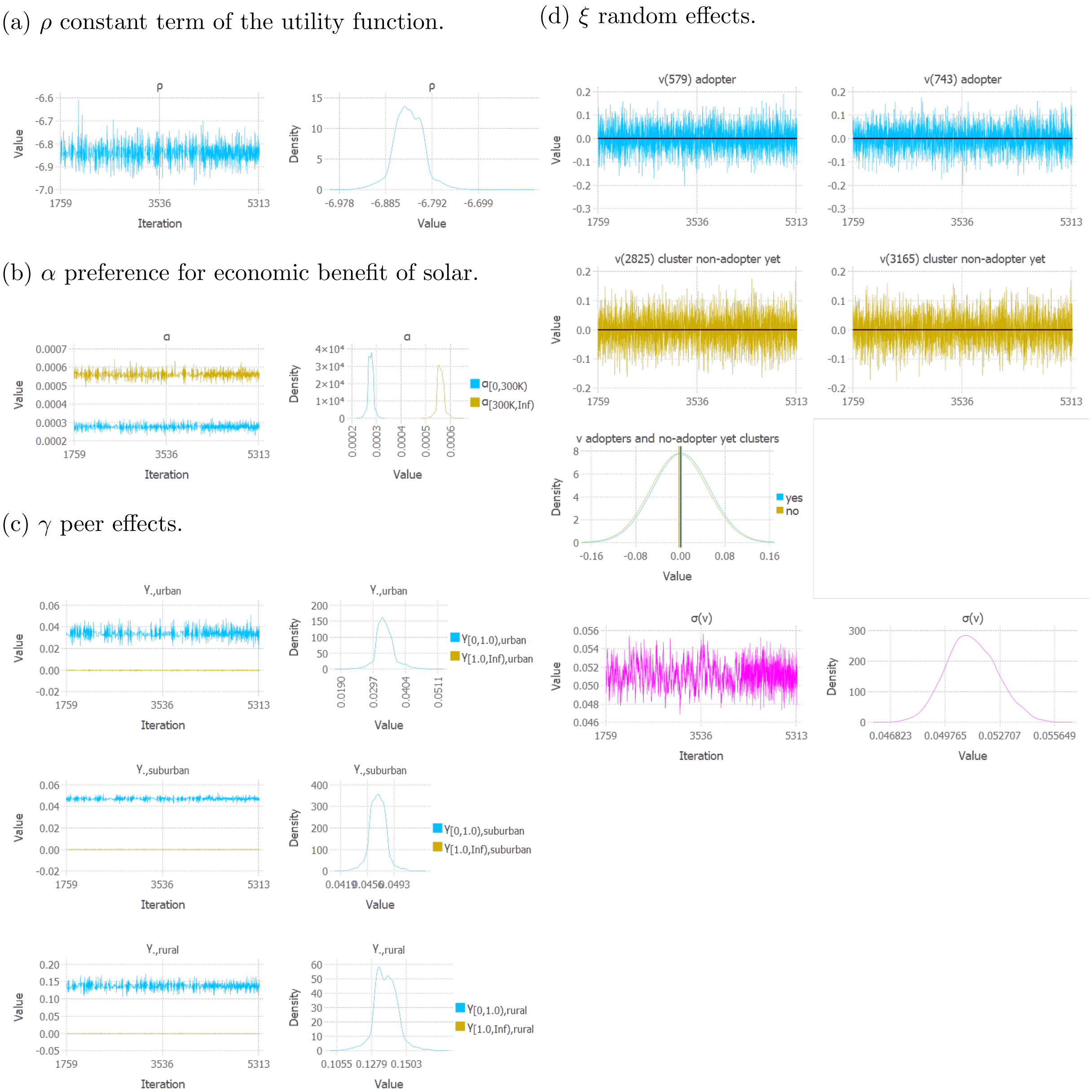
Table 3.4: Posterior mean of the best model structural parameters.

Parameter	Posterior Mean	95% Credibility Interval	Geweke p-value	Acc rate
ρ	-6.83860	$(-6.90445,-6.77342)$	0.3641	0.23
$\alpha_{[0,300\text{K})}$	0.00028	(0.00025,0.00030)	0.8520	0.26
$\alpha_{[300\text{K},\text{Inf})}$	0.00056	(0.00053,0.00059)	0.9662	0.25
$\gamma_{[0,1.0),\text{urban}}$	0.03374	(0.02751,0.04098)	0.5672	0.26
$\gamma_{[1.0,\text{Inf}),\text{urban}}$	-0.00026	$(-0.00044,-0.00008)$	0.4800	0.26
$\gamma_{[0,1.0),\text{suburban}}$	0.04702	(0.04442,0.04925)	0.9358	0.26
$\gamma_{[1.0,\text{Inf}),\text{suburban}}$	0.00005	$(-0.00006,0.00015)$	0.1299	0.26
$\gamma_{[0,1.0),\text{rural}}$	0.13716	(0.12206,0.15311)	0.9232	0.30
$\gamma_{[1.0,\text{Inf}),\text{rural}}$	-0.00033	$(-0.00053,-0.00010)$	0.9330	0.25
$\sigma(\xi)$	0.05118	(0.04874,0.05404)	0.3082	–

Note. The best model is VI-RE. DIC = 26260. Geweke tests for non-convergence of posterior mean estimates. Convergence is rejected for significant p-value. Acceptance rate is adjusted to be near 0.24. "Optimal efficiency" is achieved at an acceptance rate around that value (Roberts et al. [59]).

very small magnitude in comparison to the other parameters. That negative value is not an indication of a negative effect; rather, it shows almost zero effect. Comparing the peer effect of close installations across geographic segments, $\gamma_{[0,1.0),\text{rural}} > \gamma_{[0,1.0),\text{suburban}} > \gamma_{[0,1.0),\text{urban}}$, which indicates that rural households are influenced the most positively by each new close installation, and urban households are the least influenced.

Finally, the standard deviation of the random effect $\sigma(\xi)$ is significant (0.05118), but it has a smaller magnitude than the base utility, the economic effect, and the peer effects component of the utility function. This reflects that endogenous differences between households are significant for the decision to adopt, but with less magnitude than the other components of the utility function. The plots in Figure 3.4 show the MCMC trace and density plots of the parameters. We observe consistent convergence for all the parameters as well as the bell shape of the density plots.



(a) ρ constant term of the utility function.

(b) α preference for economic benefit of solar.

(c) γ peer effects.

(d) ξ random effects.

Figure 3.4: MCMC trace plots of best model structural parameters.

Note. Best model is Model VI-RE.

3.6.1 Comparison with Bass Model

As a benchmark, we estimated a classic Bass model [7] at the aggregate level, and we forecasted the number of adopters in out-of-sample quarters. Let A_t be the number of

(a) Quarterly number of adopters and Bass model prediction.

(b) Accumulated number of adopters and Bass model prediction.

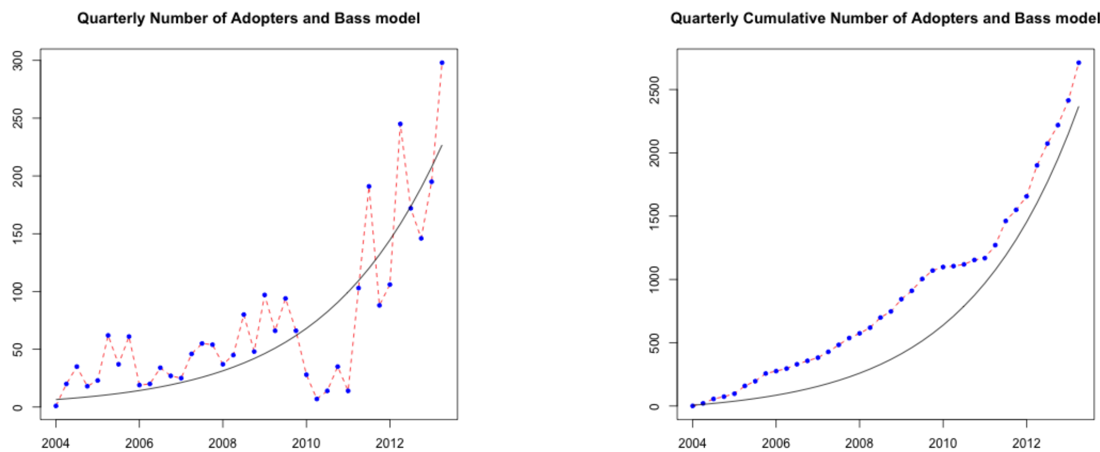


Figure 3.5: Bass model benchmark.

Note. Left panel displays the observed number of adopters per quarter and the predicted by the Bass model. Right panel displays the cumulative number. The MAPE for quarters $\mathcal{Q} := \{2011\text{-Q}1, \ldots, 2013\text{-Q}2\}$ is 83.72%.

adopters in period t, $|\mathcal{M}|$ the market size, p the coefficient of innovation, and q the coefficient of imitation. The regression to estimate is

$$A_t = |\mathcal{M}| \frac{(p+q)^2}{p} \frac{\exp(-(p+q)t)}{\left(1 + \frac{q}{p}\exp(-(p+q)t)\right)^2}.$$

We use a nonlinear least squares regression to estimate the parameters of this model. We obtain $p = 0.000139$ and $q = 0.098965$. Figure 3.5 displays both the predicted number and the actual number of adopters in quarters $\mathcal{Q} := \{2011\text{-Q}1, \ldots, 2013\text{-Q}2\}$. The MAPE on this data set is 83.72%, dramatically underperforming the best DDCM VI-RE, which has a 6.40% MAPE. The Bass model is performing poorly because the diffusion process on this data set is in a very, early stage and it does not yet attain the S-shape that the Bass equation fits.

3.7 Counterfactual Policy Analysis

Now that we have estimated the structural parameters, we can simulate the diffusion path under alternative scenarios of the exogenous state variables or the parameters that can be adjusted. In our model, a policymaker can control the ITC, the bill credits for the electricity produced (v_t), and the rebate schedule (r_t). In this essay, we focus on adjusting the rebate schedule. On the other hand, the state variable $\boldsymbol{h}_{it}$ is endogenous because it depends on the decision to adopt made by the neighbors of household i before period t. It is necessary to compute the diffusion equilibrium to obtain a transition function for $\boldsymbol{h}_{it}$ and compute the probabilities of adoption in every period for each household under a particular counterfactual.

3.7.1 Equilibrium and Simulation

Let $\bar{\boldsymbol{\theta}} = (\bar{\rho}, \bar{\boldsymbol{\alpha}}, \bar{\boldsymbol{\gamma}})$ be the posterior mean of the structural parameters. Given a specific scenario to simulate, we compute the value function and the probabilities of adoption using the fitted value iteration algorithm (see, for example, Stachurski [74] in Section 10.2.3) and an outer loop to compute the equilibrium. For each household i and period t, we estimate the value function for every possible value of the state variable $h_{it} = 1, \ldots, \hat{h}_i$, where $\hat{h}_i$ is the number of neighbors around household i. We assume a simple transition function $h_{it+1} = \lambda_i h_{it}$ [3]. We estimate λ_i at each iteration of the outer loop until the number of adopters per period converges. Initially, we set $\lambda_i = 1$ for all i.

For the last period T, the value function is computed using the Bellman operator,

[3] We experimented with more sophisticated transition functions (e.g., the Bass model), but the simulation results did not change, and the computation time increases because of the regressions on the outer loop.

iterating until convergence when the fix point of the value function is reached:

$$
\begin{aligned}
V_{iT}(c_T, h, \xi_i, \boldsymbol{\varepsilon}_i; \bar{\boldsymbol{\theta}}) &= \max\left\{\mathcal{V}_{i1T}(c_T, h, \xi_i, \boldsymbol{\varepsilon}_i; \bar{\boldsymbol{\theta}}), \mathcal{V}_{i0T}(c_T, h, \xi_i, \boldsymbol{\varepsilon}_i; \bar{\boldsymbol{\theta}})\right\} \\
V_{iT}(c_T, h, \xi_i; \bar{\boldsymbol{\theta}}) &= \exp\left(\log \mathcal{V}_{i1T}(c_T, h, \xi_i; \bar{\boldsymbol{\theta}}) + \log \mathcal{V}_{i0T}(c_T, h, \xi_i; \bar{\boldsymbol{\theta}})\right) \\
&= \exp\left(\log\left(\bar{\rho} + \bar{\alpha}_{e_i}\mathrm{NPV}_i(c_T) + \bar{\gamma}'_{g_i}h + \xi_i\right) + \log\left(\beta V_{iT}(c_T, \lambda_i h, \xi_i; \bar{\boldsymbol{\theta}})\right)\right).
\end{aligned}
$$

We estimate the fix point by computing the convergence sequence $V_i^0, V_i^1, \ldots, V_i$:

Initialization: $V_i^0 \in \mathbb{R}, k = 0$
while $||V_i^{k+1} - V_i^k|| \geq \left[\frac{(1-\beta)}{2\beta}\right]\epsilon$ **do**
  $V_i^{k+1} = \Phi V_i^k = \exp\left(\log\left(\mathcal{V}_{i1}\right) + \log\left(\beta V_i^k\right)\right)$
  $k = k + 1$
$V_i := V_i^k$.

Then, we compute the probability of adoption as

$$
\begin{aligned}
p_{iT}(c_T, h, \xi_i; \bar{\boldsymbol{\theta}}) &= \frac{1}{1 + \exp\left(\mathcal{V}_{i0T}(c_T, h, \xi_i; \bar{\boldsymbol{\theta}}) - \mathcal{V}_{i1T}(c_T, h, \xi_i; \bar{\boldsymbol{\theta}})\right)} \\
&= \frac{1}{1 + \exp\left(\beta V_{iT}(c_T, \lambda_i h, \xi_i; \bar{\boldsymbol{\theta}}) - \mathcal{V}_{i1T}(c_T, h, \xi_i; \bar{\boldsymbol{\theta}})\right)}.
\end{aligned}
$$

For period $t < T$, the value function can be computed using the backward recursion:

$$
\begin{aligned}
V_{it}(c_t, h, \xi_i, \boldsymbol{\varepsilon}_i; \bar{\boldsymbol{\theta}}) &= \max\left\{\mathcal{V}_{i1t}(c_t, h, \xi_i, \boldsymbol{\varepsilon}_i; \bar{\boldsymbol{\theta}}), \mathcal{V}_{i0t}(c_t, h, \xi_i, \boldsymbol{\varepsilon}_i; \bar{\boldsymbol{\theta}})\right\} \\
V_{it}(c_t, h, \xi_i; \bar{\boldsymbol{\theta}}) &= \exp\left(\log \mathcal{V}_{i1t}(c_t, h, \xi_i; \bar{\boldsymbol{\theta}}) + \log \mathcal{V}_{i0t}(c_t, h, \xi_i; \bar{\boldsymbol{\theta}})\right) \\
&= \exp\left(\log\left(\bar{\rho} + \bar{\alpha}_{e_i}\mathrm{NPV}_i(c_t) + \bar{\gamma}'_{g_i}h + \xi_i\right) + \log\left(\beta V_{it+1}(c_{t+1}, \lambda_i h, \xi_i; \bar{\boldsymbol{\theta}})\right)\right),
\end{aligned}
$$

and the probability of adoption is

$$
\begin{aligned}
p_{it}(c_t, h, \xi_i; \bar{\boldsymbol{\theta}}) &= \frac{1}{1 + \exp\left(\mathcal{V}_{i0t}(c_t, h, \xi_i; \bar{\boldsymbol{\theta}}) - \mathcal{V}_{i1t}(c_t, h, \xi_i; \bar{\boldsymbol{\theta}})\right)} \\
&= \frac{1}{1 + \exp\left(\beta V_{it+1}(c_{t+1}, \lambda_i h, \xi_i; \bar{\boldsymbol{\theta}}) - \mathcal{V}_{i1t}(c_t, h, \xi_i; \bar{\boldsymbol{\theta}})\right)}.
\end{aligned} \tag{3.8}
$$

By using the probabilities of adoption per household, it is possible to simulate and compute statistics over many simulations. However, that process is computationally intensive. Alternatively, we compute the expected number of adopters directly using the probabilities of adoption.

We must treat two caveats carefully. The first caveat is the use of clusters. Let us recall that a *cluster* is a group of households from the same block group that shares similar characteristics (i.e., home market value, size of the house, tree cover, and irradiance received). In the estimation process described in Section 3.5, the clusters are treated as individual households, except in the likelihood function. The likelihood of each cluster is weighted by the number of households that belong to it. Equivalently, in the simulation, we degroup each cluster by the number of households that belong to that cluster. Next, we assign a probability of adoption to each household equal to the probability of the corresponding cluster. Then, we simulate the decision-making process of each household per period, considering the installation net cost and the adopters from previous periods.

As we stated earlier in this section, the second caveat is that the state variable of neighbors that have adopted previously must be found in equilibrium. For this purpose, we iterate the simulation until the number of adopters converges using the number of adopters from the previous iteration to regress $h_{it+1} = \lambda_i h_{it}$. The algorithm in Table 3.5 describes the steps of the simulation. In this way, we assume that households project the equilibrium number of adopters, and households do not consider the actions of potential adopters in the actual period, as in the concept of oblivious equilibrium of Weintraub et al. [82, 83]. Figure 3.6 shows an example of how the equilibrium is reached via the outer loop.

3.7.2 Policy Analysis

We present two counterfactuals in this essay. We alter the net cost in the first counterfactual, and we alter the rebate in the second. We start with the ideal hypothetical situation of changing the net cost, because that variable plays a direct role on the utility function defined in Equation (3.1). However, in the real world, a policymaker can adjust only the rebate. Thus, we present that analysis in the altered rebate counterfactual below.

Table 3.5: Algorithm expected number of adopters in equilibrium.

Input : Schedule of net cost $c_t, t = 1, \dots, T$.
Output : Expected number of adopters for periods $t = 1, \dots, T$.
Initialization: No household have adopted in period $t = 1$. Iteration $k \leftarrow 1$. $\lambda_i = 1$ for each household i
while *Number of adopters do not converge* **do**
  for $i \in$ *Households* **do**
    household i adopt in $t = 1$ with probability $P_{i1} = p_i(c_t, 0, \xi_i; \bar{\boldsymbol{\theta}})$
  for $i \in$ *Households* **do** `// it is necesary to finish previous loop`
    $h_{i1}^{(k)}$ = expected number of adopters at a 1-mile radius of household i at the end of period $t = 1$
  $A_1^{(k)}$ = expected number of adopters in period 1 of iteration k
  for $t \in 2$ **to** T **do**
    for $i \in$ *Households that have not yet adopted* **do**
      household i adopt in period t with probability $P_{it} = p_i(c_t, h_{it-1}^{(k)}, \xi_i; \bar{\boldsymbol{\theta}}) * (1 - \sum_{\tau=1}^{t-1} P_{i\tau})$
    for $i \in$ *Households* **do**
      $h_{it}^{(k)}$ = expected number of adopters at a 1-mile radius of household i at the end of period t
    $A_t^{(k)}$ = expected number of adopters in period t of iteration k
  for $i \in$ *Households* **do**
    Estimate linear regression to update λ_i: $h_{it+1}^{(k)} = \lambda_i h_{it}^{(k)}$
  $k \leftarrow k + 1$;

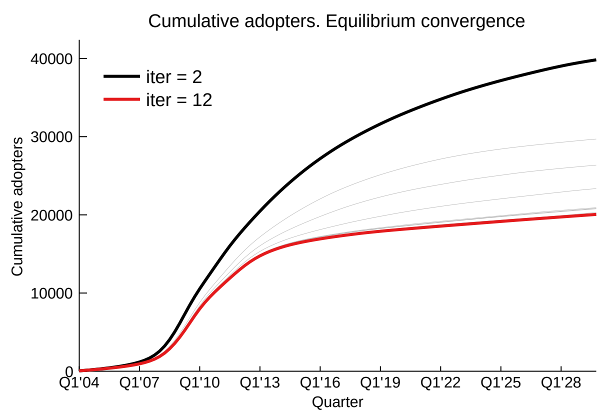


Figure 3.6: Example of equilibrium reached via outer loop.

3.7.2.1 Net Cost Counterfactual

We compare two scenarios to the base case of a smooth average net cost of installation. The metrics that we evaluated are (1) the expected cumulative number of adopters by quarter and (2) the expected budget spent on rebates by quarter. We designed the counterfactual scenarios as follows. Assume the hypothetical scenario in which the government, which provides the rebates, announces a scheduled net cost of installation $\tilde{\boldsymbol{c}} = \{\tilde{c}_1, \dots, \tilde{c}_T\}$ in

\$/kW for the period of evaluation, Q1-2004 to Q3-2013. Recall that the net cost per kW, c_t, is equal to the price of installation, p_t, minus the rebate incentive, r_t, discounted by the federal incentive tax credit, ITC_t, which was 30% during the period of evaluation, $c_t = (p_t - r_t)(1 - \text{ITC}_t)$. Therefore, in this hypothetical situation, the government adjusts the rebate to satisfy the announced schedule as $\tilde{r}_t = p_t - (\tilde{c}_t/(1 - \text{ITC}_t))$.

The net cost of installation scenarios are (i) constant net cost; (ii) stepwise increasing net cost, and (iii) stepwise decreasing net cost. We set the constant net cost equal to the average real net cost, \$1,640. The stepwise curves are constructed using a factor in $\{0.2, 0.3\}$ such that minimum cost = mean cost×(1−factor), maximum cost = mean cost×(1+factor), and using six evenly distributed steps from minimum cost to maximum cost. For example, for factor 0.3 the minimum cost is $\$1,640 \times (1 - 0.3) = \$1,148$, and the maximum cost is $\$1,640 \times (1 + 0.3) = \$2,132$.

In Figure 3.7, Plots (a1) and (b1) present the net cost curves per quarter. We name the scenario in which the factor is 0.2 the narrow range scenario, and the wide range scenario has a factor of 0.3. These same plots show the constant net cost scenario and the smoothed base case scenario. Plots (a2) and (b2) show the rebate schedules that resulted from adjusting to the net cost. Notice that the scenario rebate schedules are not far from the actual rebate schedule. Plots (a3) and (b3) present the accumulated number of adopters, and Plots (a4) and (b4) present the accumulated budget spent.

Let us focus our attention on the wide range scenario. Plot (b3) shows that at the end of Q3-2013, the actual number of adopters is 2,722, and the predicted number of adopters with the smooth base case is 2,877. With constant net cost, the number of adopters is 2,883, a 0.21% increase with respect to the base case. With decreasing net cost, the number of adopters is 2,932, a 1.91% increase with respect to the base case; and with increasing net cost the number of adopters is 3,672, a raise of 27.63% with respect to the base case.

Observing the accumulated budget spent on rebates in Plot (b4), we see that the

Table 3.6: Percent difference in number of adopters and budget spent for net cost scenarios.

Net Cost Scenario	Adopters % difference	Budget Spent % difference
Constant	0.21%	−2.65%
Decreasing Narrow	0.31%	−10.38%
Decreasing Wide	1.91%	−8.46%
Increasing Narrow	11.51%	26.22%
Increasing Wide	27.63%	55.81%

Note. The base case predicts 2,877 adopters and a total budget spent of $24.94 million at the end of Q2-2013. We observe that constant and stepwise decreasing net cost scenarios incentivize more adopters at a lower total budget. Stepwise increasing net cost incentivizes many more adopters but with a higher total budget.

actual budget spent is $26.24 million and the predicted budget spent with the smooth base case scenario is $24.94 million. In the constant net cost scenario, the budget spent is $24.28 million (i.e, 2.65% less than the base case). With decreasing net cost the budget spent is $22.83 million (i.e, 8.46% less than the base case). And with increasing net cost the budget spent is $38.86 dollars (i.e, 55.81% more than the base case). Table 3.6 summarizes these results. Compared to the base case, a constant and stepwise decreasing net cost of installation are Pareto superior in the sense that these scenarios incentivize more adopters than the base case at a lower total budget spent on rebates. A stepwise increasing net cost incentivizes 27.6% more adopters than the base case, but with an average increased rebate of 22% per adopter.

As we commented earlier on this section, these rebate schedules are comparable to the actual rebate schedule. Relatively small changes in these curves have substantial consequences on the number of adopters and the budget spent. Our framework can compute the consequences of applying particular scenarios beforehand. Using our framework, we can design and evaluate more sophisticated counterfactuals, such as active rebates that target specific areas of the city in particular periods to kick start peer effects.

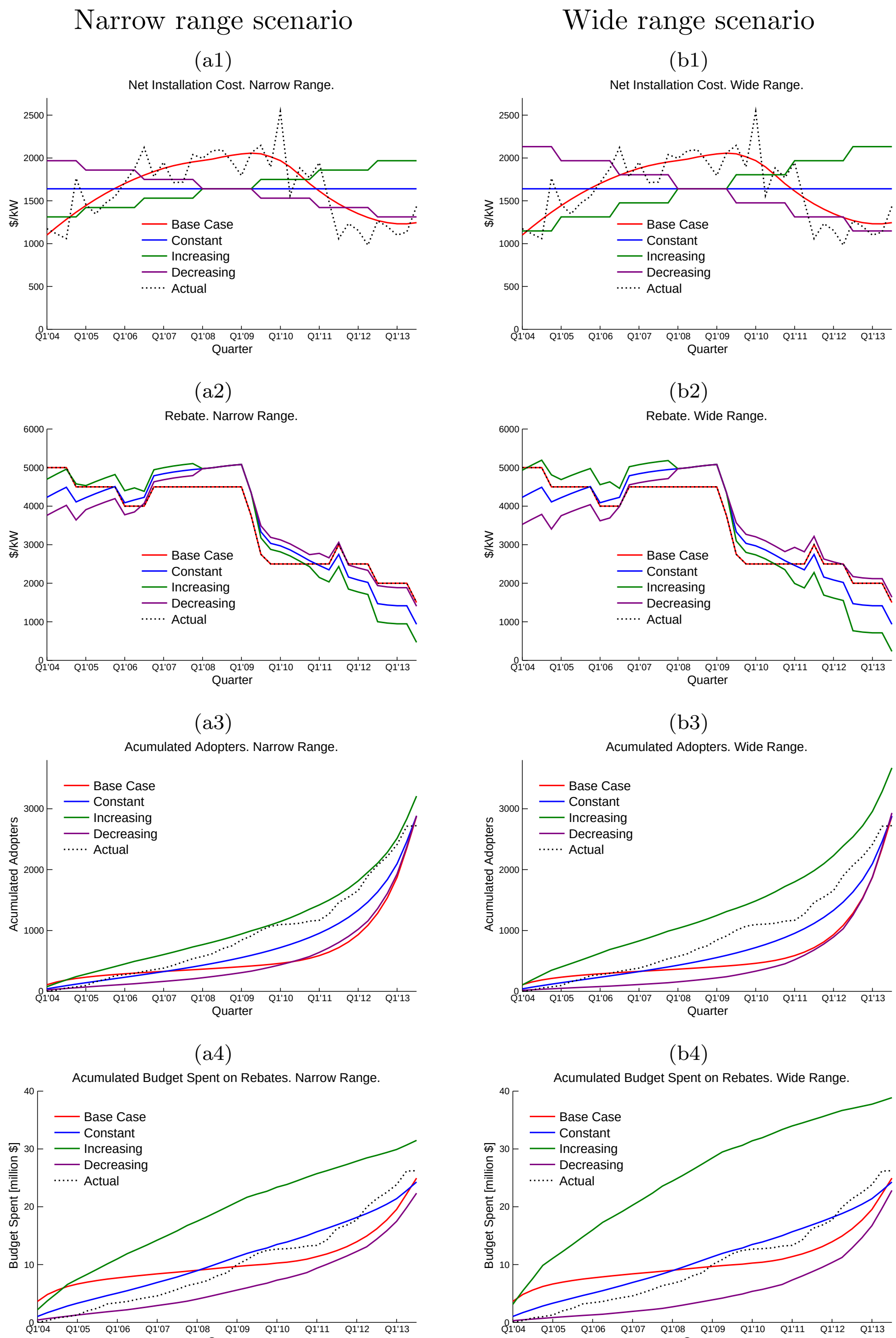


Figure 3.7: Net cost counterfactual. Net cost of installation scenarios, equilibrium adoption, and budget spent on rebates.

Note: The net cost of installation per kW, c_t, is equal to the installation price charged by the installer, p_t, minus the rebate incentive, r_t, discounted by the federal incentive tax credit ITC_t (which was 30% during the period of evaluation). Hence, $c_t = (p_t - r_t)(1 - \text{ITC}_t)$.

3.7.2.2 Rebate Counterfactual

In this counterfactual, we alter the rebate and simulate farther into the future until Q4-29. The price history available is from Q1-04 until Q4-14. We fit a smooth curve using locally estimated scatterplot smoothing (LOESS, Cleveland et al. [24]). After the Q4-14 quarter, we assume a constant price equal to the price in Q4-14, which is $2.50 per watt after ITC. See Figure 3.8 (a1) curve Smooth price.

We compare three rebate scenarios. In Scenario S1, the rebate starts at $3.55 per watt and decreases at a rate such that the difference between the price and rebate (i.e., the net cost) is constant at $1.64 per watt (Figure 3.8 (a1)). In Scenario S2, the rebate also starts at $3.55 per watt and decreases at a rate such that the net cost is constant at $1.64 per watt. But, in quarter Q1-18, the rebate drops to zero and the net cost increases to $2.491 per watt (Figure 3.8 (b1)). The difference between Scenarios S1 and S2 is the drop of the rebate to zero in a given quarter. In Scenario S3, the rebate starts at $3.22 per watt and decreases at a rate such that the net cost is constant at $1.968 per watt. Then, in quarter Q1-18, the rebate drops to zero and the net cost increases to $2.491 per watt (Figure 3.8 (c1)). The only difference between Scenarios S2 and S3 is the starting value of the rebate.

For these three scenarios, we simulate the diffusion process, and we evaluate the following four metrics. Metric M1 is the expected cumulative number of adopters. Metric M2 is the expected cumulative megawatts installed. Metric M3 is the expected cumulative budget spent on rebates. Metric M4 is the expected budget spent per megawatt installed.

Comparing Scenario S1 and S2 by Metric M1, we observe that the cumulative number of adopters is higher in Scenario S2 . This may be a surprising result, given that in Scenario S2, the rebate is lower or equal to the rebate in Scenario S1 in any given quarter. For a forward-looking household, it is rational to adopt earlier because the rebate will drop in the future. As a consequence, the peer effects arise sooner than they do in Scenario S1, motivating even more households to adopt. Subsequently, for Scenario S2, the cumulative

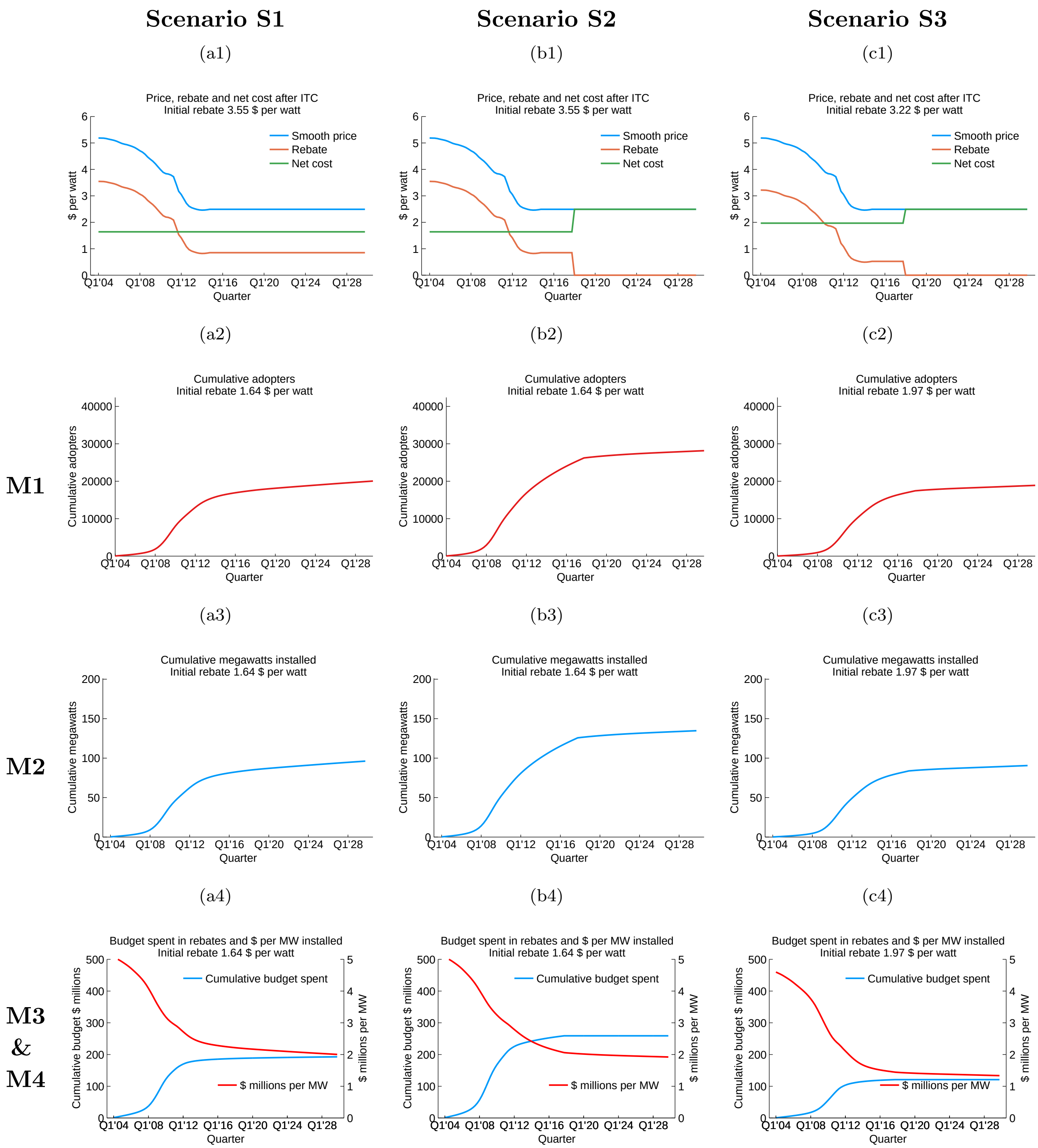


Figure 3.8: Rebate counterfactual. Net cost of installation scenarios, equilibrium adoption, megawatts installed and budget spent on rebates.

megawatts (Metric M2) is higher, and cumulative budget spent (Metric M3) is also higher. Interestingly, the ratio between these two metrics—the budget spent per megawatt installed (Metric M4)— is lower in Scenario S2. When we compare Scenarios S2 and S3, we observe a difference in the net cost of $(\$1.968-\$1.64)/\$1.64\times 100 = 20\%$ during the first 14 years (Q1-04 to Q1-18), which is high enough to deter many households from adopting. The diffusion process in the long term is almost 30% lower for Scenario S3 when compared to Scenario S2.

3.8 Conclusions

Solar energy will be an essential component of the future energy portfolio. It is a clean and renewable fuel for generating electricity. However, we are still in an early phase of the diffusion curve. In order to reach higher levels of penetration, public and private interventions are necessary to incentivize adoption. To provide insights for policy design, we developed a diffusion model that takes into consideration the dynamics of decision-making, economic analysis, and the influence of previous adopters. We have brought together a unique data set of PV adopters, which allows us to estimate the model at a micro-level of resolution. We characterized the adopters by income and urbanization level, estimating the predisposition of potential adopters to install PV given the economic considerations and the peer effects.

Our estimates reflect that mid- to high-level income households are more keen to adopt PV, and PV installations that occur nearby have a more significant influence over potential adopters. The novelty of this research lies in its quantification of the effects of the dynamic environment on household preferences. Our counterfactual analysis demonstrates that fixed or decreasing net cost curves incentivize more adopters than the base case at a lower average rebate per adopter, while increasing the net cost incentivizes many more adopters but at a higher average rebate per adopter. Our framework can help policymakers design an optimal rebate schedule that maximizes PV adoption for a given budget level.

Chapter 4

Conclusions and Future Research

This dissertation consists of two essays that seek to predict and influence consumer demand. In the first essay, we designed and implemented an end-to-end solution to scheduling advertising on linear television. Leading networks in the United States and in India use our approach, increasing their ad revenue by 3% to 5%. In the second essay, we developed a structural econometric model to predict solar panel adoption and to design policies to incentivize their diffusion.

The first essay addresses the problem of allocating ad videos during commercial breaks. The contracts between the network and advertisers specify that the ads must be shown to a specific number of viewers from a targeted demographic. An ad can be aired as many times as necessary until the intended number of viewers is reached. Moreover, the advertisers and the network both place a series of constraints on the ads.

We formulate the problem as one large-scale stochastic programming model that impossible to solve in practice. We produce asymptotic results that allow us to use a ratings point forecast to simplify the stochastic problem into a deterministic model. Still, it is not possible to solve the complete mathematical programming model with all the associated constraints in the short amount of time that is available each day (i.e., less than 15 minutes). For that purpose, we separate the problem into three stages that reduce the computational time to a few minutes obtaining schedules that are close to the optimal solution. Also, we designed time series algorithms that produce accurate ratings point forecast.

This essay establishes several avenues for future research. The first avenue is the development of a more sophisticated approach to address the uncertainty of the ratings. One possibility that we have explored is robust optimization that exploits the correlation between the ratings of different commercial breaks. In practice, the new formulation only needs a few more constraints that does not add more complexity to the problem.

A second line of research we plan to explore is of machine learning methods and structural estimation methods to improve the accuracy of the ratings forecasts. For a different application, scheduling programs instead of ads, we have implemented gradient boosting trees algorithms to forecast ratings, thus improving accuracy relative to the time series model described in Chapter 2. We plan to interpret the resulting trees specify the most important features of ratings prediction. Understanding these features will improve the allocation of ads during commercial breaks to target certain demographics.

The disaggregated data that we use to compute ratings are the minute-by-minute viewership path of each of the household members that compose the Nielsen panel (about 80,000 households). This viewership data set is joined with demographic and geographic information. We plan to analyze this data set using a structural model inspired by our second essay. The research question is how to predict future viewership per household and how this viewership would change if ads and programs air in a different order. Consequently, networks could gain a better understanding of their viewers and offer a better marketing mix to advertisers.

Methodologically, no structural estimation method can handle a problem of this size. A conservative estimation of the panel size is 80,000 agents $\times$ 24 hours per day $\times$ 365 days per year $\times$ 3 years of historical data. This equals 2,102.4 million rows of decisions with approximately 200 covariates (agents and shows characteristics). That is a very large dynamic discrete choice (DDC) estimation problem. We must design an estimation method that can address such a massive problem.

Our second essay investigated the diffusion of residential solar panel (PV) systems.

PV adoption in the United States is proliferating. According to the Solar Energy Industries Association PV systems installations showed an average annual growth rate of 50% during the previous decade. This growth has been propelled by federal and local incentives and the decline of installation costs, which was approximately 70% during the last decade. However, these two factors alone cannot explain such exponential growth. We expand on a decade of research produced by the Energy Systems Transformation Research Group of The University of Texas at Austin as we develop a DDC model that considers the economic and social factors that influence the adoption decision. We design a Bayesian estimation method that allows us to handle heterogeneity gracefully. Our structural parameter estimates inform the relative importance of the economic factors and social influence on the utility function of households. Our estimates also allow us to evaluate counterfactual incentive scenarios so that a policymaker can design the optimal schedule of incentives to maximize a given metric (e.g., the cumulative number of adopters, pollution savings).

We plan to compare our empirical results with the theoretical work on technology adoption of Ulu and Smith [76] and Smith and Ulu [69, 70], who studied the optimality conditions for the time of adoption under functional assumptions. This comparison can lead to a more refined utility function in future empirical studies. We also plan to investigate new estimation methods that incorporate the discount factor as a structural parameter. We followed the traditional approach to guess a reasonable discount factor. But, perhaps, this factor could be estimated from the data as Abbring and Daljord [1] discuss.

On the methodological front, we intend to extend our estimation method to handle a more agents, decisions, covariates, and time periods to model other types of dynamic decision problems in which the evaluation of counterfactual policies have a strategic value. We plan to combine machine learning and structural econometric techniques to develop faster and more precise estimation methods for dynamic discrete choice models.

Appendices

Appendix A

Scheduling Advertising on Television

A.1 Disaggregated Results for August of a Sample Year

This appendix presents the disaggregated data that is used to construct the summary statistics presented in the essay. Data in Tables A.1, A.2, A.3, A.4 and A.5 are the disaggregate data of essay's Tables 6, 7, 8, 9 and 10 respectively.

Table A.1: Statistics of instances august sample year.

Instance-Day	# Breaks $\lvert\mathcal{B}\rvert$	# Demos $\lvert\mathcal{Q}\rvert$	# Advertisers $\lvert\mathcal{A}\rvert$	# Brands $\lvert\mathcal{K}\rvert$	# Deals $\lvert\mathcal{D}\rvert$	# Spots $\lvert\mathcal{U}\rvert$	# PCat. $\lvert\mathcal{P}\rvert$
8/1	132	16	69	138	98	622	255
8/2	129	18	79	160	121	624	255
8/3	126	18	82	166	133	663	255
8/4	129	19	93	194	155	728	255
8/5	133	18	90	181	152	715	255
8/6	127	17	83	173	109	746	255
8/7	135	18	85	160	118	786	255
8/8	135	18	76	135	100	765	255
8/9	128	16	86	174	129	655	255
8/10	123	14	88	172	138	683	255
8/11	128	14	86	173	127	669	255
8/12	129	14	95	176	152	822	255
8/13	132	12	84	158	113	858	255
8/14	132	13	75	136	100	697	255
8/15	137	13	80	147	106	756	255
8/16	171	18	88	165	128	579	255
8/17	167	18	85	177	132	638	255
8/18	169	16	76	167	117	636	255
8/19	171	15	84	160	126	607	255
8/20	172	16	81	153	112	680	255
8/21	173	16	81	159	113	635	255
8/22	176	15	73	121	93	464	255
8/23	165	17	83	126	129	582	255
8/24	161	17	86	186	143	657	255
8/25	168	16	81	170	129	625	255
8/26	170	18	91	183	144	621	255
8/27	173	15	85	174	124	745	255
8/28	173	16	85	159	115	657	255
8/29	170	16	72	134	97	505	255
8/30	156	21	98	210	162	717	255
8/31	168	20	88	185	142	700	255

Note. Stdev: standard deviation. CV: coefficient of variation (Stdev/mean). PCat: product categories.

Table A.2: Stage 2 problems size, and MIP gap.

Instance-Day	# Binary vars.	# Continuous vars.	# Constraints	MIP gap [%]
8/1	7,248	11,187	20,445	0.82
8/2	8,875	10,502	18,579	1.01
8/3	7,495	9,955	18,022	0.03
8/4	9,865	12,034	21,419	0.02
8/5	11,064	11,971	21,562	0.06
8/6	18,033	18,302	32,181	0.79
8/7	8,560	12,889	23,007	0.62
8/8	10,274	13,947	25,103	0.83
8/9	6,623	9,086	16,451	0.76
8/10	5,428	7,984	14,693	0.74
8/11	9,224	11,723	20,847	0.04
8/12	7,429	10,498	19,016	0.96
8/13	6,845	10,230	19,135	0.99
8/14	7,378	10,322	19,130	1.00
8/15	12,968	16,241	29,659	0.99
8/16	5,106	7,942	14,159	0.28
8/17	12,517	14,297	24,860	0.78
8/18	8,532	10,894	20,022	0.99
8/19	10,669	12,101	21,686	0.08
8/20	21,332	21,495	37,548	1.37
8/21	17,095	17,955	31,709	0.03
8/22	4,176	6,470	12,615	0.98
8/23	2,188	4,728	9,335	0.21
8/24	12,350	14,600	25,657	0.26
8/25	24,694	23,538	39,791	0.12
8/26	20,861	21,196	36,292	0.24
8/27	37,418	32,206	55,563	0.43
8/28	23,462	22,058	38,624	0.32
8/29	6,277	9,256	16,555	0.70
8/30	8,980	11,744	20,954	0.78
8/31	8,512	10,822	19,610	0.48

Note. MIP gap = ((MIP objective function - best LP bound objective function)/best LP bound objective function)$\times$ 100.

Table A.3: Weighted average ratings (M1) and value by day based on forecast ratings.

	M1			Value		
Day	ORG	OPT	gain [%]	ORG [$]	OPT [$]	gain [%]
8/1	115,915,000	117,609,000	1.46	1,565,770	1,589,100	1.49
8/2	84,954,200	86,191,100	1.46	1,190,980	1,207,240	1.37
8/3	107,126,000	108,123,000	0.93	1,962,410	1,979,820	0.89
8/4	99,418,400	100,947,000	1.54	1,541,040	1,563,640	1.47
8/5	75,768,000	77,270,100	1.98	1,109,250	1,131,060	1.97
8/6	73,366,900	75,414,800	2.79	1,192,450	1,221,150	2.41
8/7	94,228,400	95,529,900	1.38	1,503,240	1,519,120	1.06
8/8	98,232,000	100,735,000	2.55	1,691,920	1,723,760	1.88
8/9	78,514,900	79,433,600	1.17	1,289,620	1,302,440	0.99
8/10	106,816,000	107,852,000	0.97	1,773,440	1,793,060	1.11
8/11	81,675,700	82,654,400	1.20	1,347,240	1,359,650	0.92
8/12	76,324,400	77,427,200	1.44	1,320,000	1,339,410	1.47
8/13	67,917,900	68,933,900	1.50	1,147,950	1,165,090	1.49
8/14	87,648,000	90,476,000	3.23	1,445,980	1,489,430	3.00
8/15	83,754,100	85,588,900	2.19	1,356,760	1,381,780	1.84
8/16	76,467,100	77,322,700	1.12	1,223,040	1,237,260	1.16
8/17	98,908,900	101,478,000	2.60	1,844,080	1,894,520	2.74
8/18	86,515,300	88,296,800	2.06	1,479,870	1,502,280	1.51
8/19	78,213,600	79,785,100	2.01	1,412,880	1,437,550	1.75
8/20	66,027,300	69,624,400	5.45	1,084,500	1,142,620	5.36
8/21	76,425,900	78,742,200	3.03	1,223,300	1,258,260	2.86
8/22	66,274,700	67,104,600	1.25	1,105,080	1,118,590	1.22
8/23	78,573,000	78,941,700	0.47	1,274,720	1,279,520	0.38
8/24	111,999,000	114,717,000	2.43	2,028,790	2,088,610	2.95
8/25	89,990,800	95,485,200	6.11	1,520,670	1,606,590	5.65
8/26	97,527,000	100,115,000	2.65	1,805,710	1,843,780	2.11
8/27	89,293,200	93,426,700	4.63	1,459,230	1,525,290	4.53
8/28	84,572,300	87,770,200	3.78	1,269,710	1,324,050	4.28
8/29	66,996,400	68,641,000	2.45	1,028,270	1,048,840	2.00
8/30	88,940,200	90,928,100	2.24	1,378,660	1,406,970	2.05
8/31	114,625,000	115,171,000	0.48	2,033,440	2,045,400	0.59

Note. ORG: original schedule. OPT: optimized schedule. gain = ((OPT-ORG)/ORG)$\times$ 100. We use a random scaling factor to maintain confidentiality, but the order of magnitudes correspond to the true values.

Table A.4: Sum of penalties by day.

Penalties M2+M3+M4+M5		
ORG	OPT	ORG/OPT
1,609,370	0	-
6,093,750	556,003	10.96
5,387,620	142,321	37.86
9,428,160	226,893	41.55
2,687,760	0	-
11,860,200	798,251	14.86
2,399,360	14,316	167.61
3,546,980	0	-
4,968,830	186,559	26.63
6,248,010	332,281	18.80
5,823,790	198,591	29.33
3,373,700	0	-
2,462,360	219,410	11.22
4,063,760	234,747	17.31
4,324,200	215,280	20.09
3,749,290	0	-
2,438,150	44,884	54.32
2,538,320	266,903	9.51
2,385,700	0	-
2,496,930	0	-
3,313,990	68,886	48.11
741,317	0	-
2,863,230	0	-
3,418,260	151,616	22.55
2,680,790	0	-
1,791,720	0	-
3,753,100	0	-
3,712,180	0	-
982,697	0	-
3,153,530	19,800	159.27
1,484,170	0	-

Note. ORG: original schedule. OPT: optimized schedule. We use a random scaling factor to maintain confidentiality, but the order of magnitudes correspond to the true values.

Table A.5: Allocated ratings and value by day based on actual ratings.

Day	M1 ORG	M1 OPT	M1 gain [%]	Value ORG [$]	Value OPT [$]	Value gain [%]
8/1	120,469,000	121,577,000	0.92	1,651,300	1,674,730	1.42
8/2	98,553,200	99,305,600	0.76	1,375,570	1,388,250	0.92
8/3	109,442,000	110,609,000	1.07	2,173,350	2,196,640	1.07
8/4	123,849,000	125,280,000	1.16	2,028,000	2,050,400	1.10
8/5	159,841,000	161,165,000	0.83	2,817,310	2,832,800	0.55
8/6	112,041,000	113,524,000	1.32	1,820,170	1,837,830	0.97
8/7	124,519,000	125,322,000	0.64	1,987,650	1,996,780	0.46
8/8	92,648,400	93,947,800	1.40	1,558,510	1,576,390	1.15
8/9	85,535,900	86,291,600	0.88	1,440,120	1,456,780	1.16
8/10	108,325,000	109,629,000	1.20	1,884,040	1,929,940	2.44
8/11	92,348,500	92,280,000	-0.07	1,596,280	1,590,920	-0.34
8/12	135,347,000	137,014,000	1.23	2,661,940	2,693,500	1.19
8/13	90,576,700	91,241,600	0.73	1,547,220	1,552,710	0.35
8/14	89,458,100	91,624,100	2.42	1,483,750	1,520,410	2.47
8/15	84,433,800	84,591,700	0.19	1,377,600	1,380,390	0.20
8/16	94,496,100	94,551,500	0.06	1,464,360	1,473,050	0.59
8/17	121,118,000	122,386,000	1.05	2,317,580	2,354,000	1.57
8/18	109,291,000	111,762,000	2.26	1,940,750	1,978,660	1.95
8/19	140,016,000	140,878,000	0.62	2,754,810	2,781,890	0.98
8/20	89,270,100	94,830,900	6.23	1,459,720	1,542,500	5.67
8/21	71,132,400	73,196,100	2.90	1,166,750	1,191,750	2.14
8/22	71,899,700	72,694,700	1.11	1,211,100	1,225,200	1.16
8/23	93,019,200	93,484,400	0.50	1,559,620	1,567,760	0.52
8/24	94,542,800	94,859,100	0.33	1,739,730	1,748,140	0.48
8/25	95,303,100	101,850,000	6.87	1,726,650	1,821,010	5.46
8/26	151,335,000	154,379,000	2.01	3,118,450	3,141,130	0.73
8/27	105,478,000	106,425,000	0.90	1,697,640	1,708,100	0.62
8/28	98,512,300	102,000,000	3.54	1,522,060	1,578,000	3.68
8/29	74,459,300	75,820,100	1.83	1,159,240	1,165,810	0.57
8/30	110,574,000	112,516,000	1.76	1,827,100	1,836,320	0.50
8/31	139,638,000	139,636,000	0.00	2,530,300	2,527,880	-0.10

Note. ORG: original schedule. OPT: optimized schedule. gain = ((OPT-ORG)/ORG)$\times$ 100. We use a random scaling factor to maintain confidentiality, but the order of magnitudes correspond to the true values. Actual ratings are obtained after the spots are aired.

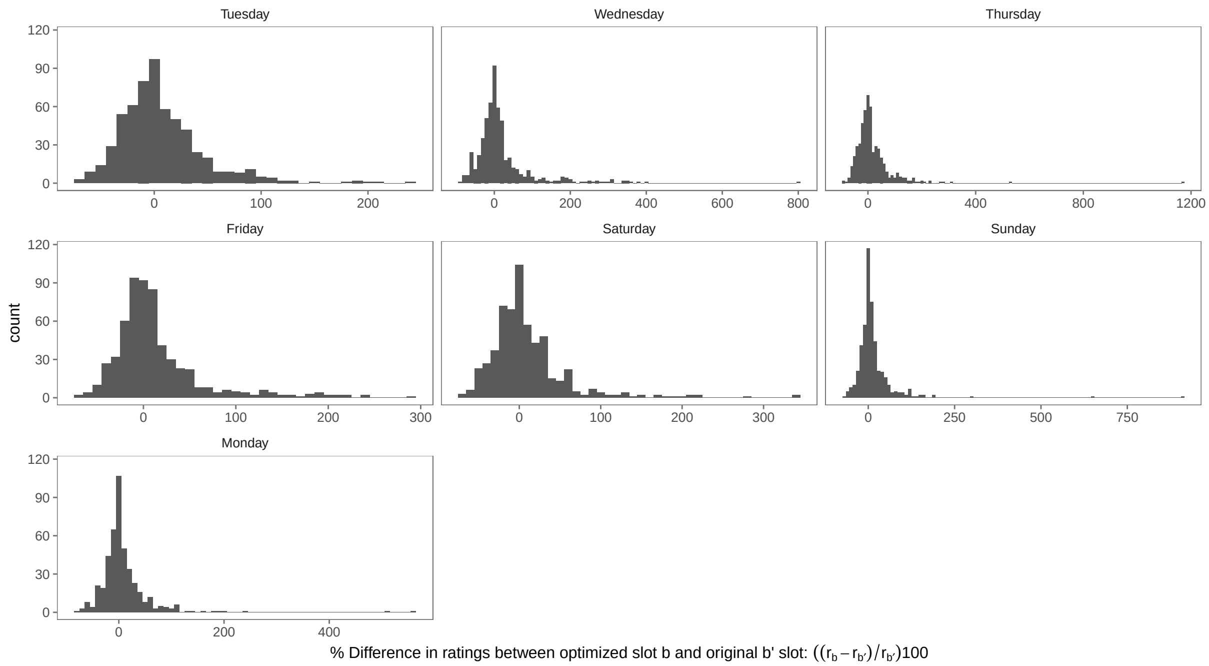


Figure A.1: % Lift histogram per move, per day of the week august of sample year

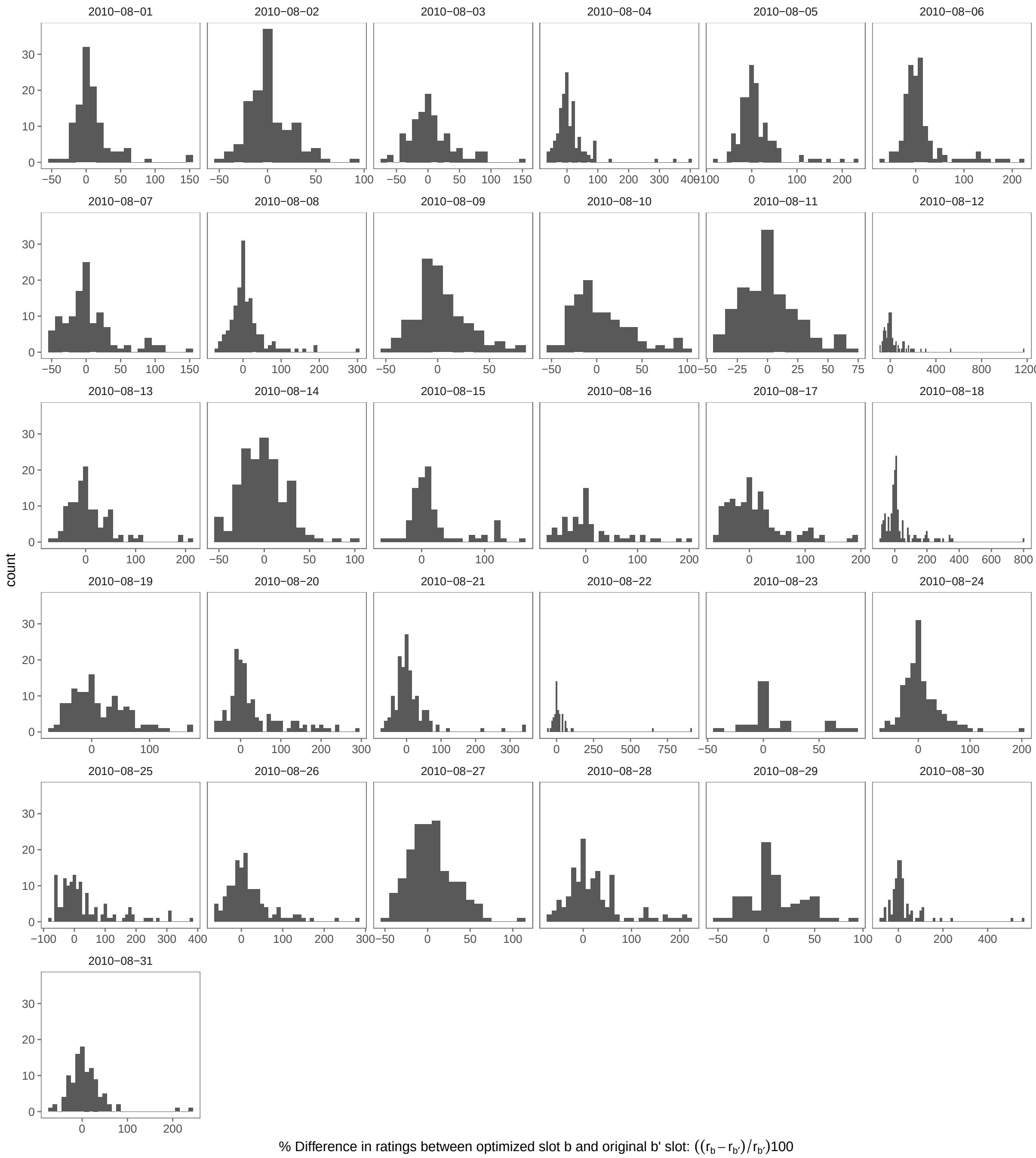
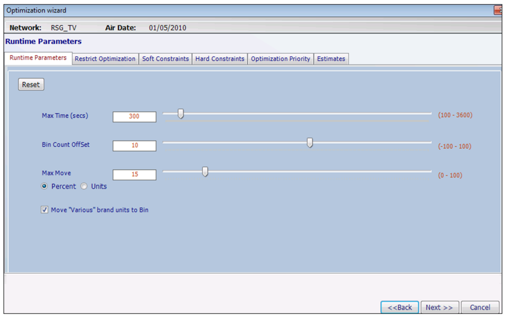


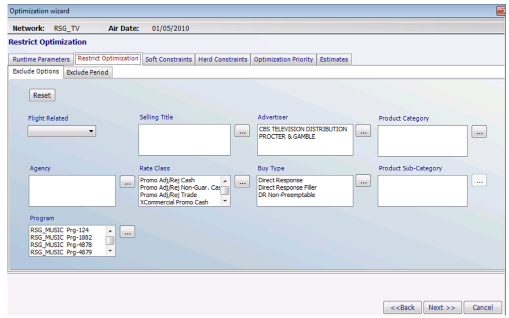
Figure A.2: % Lift histogram per move, per day of August of sample year

A.2 Application Screens

(a) Setting of general parameters

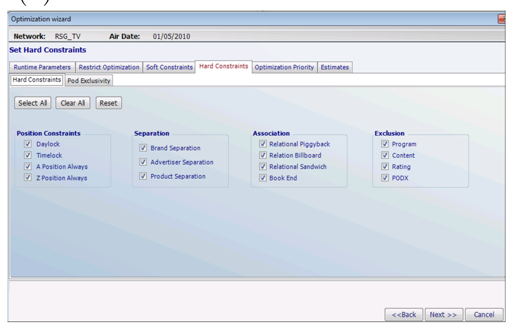


(b) Definition of instance

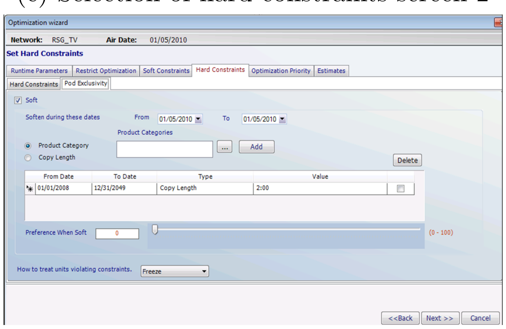


(c) Setting of soft constraints

(d) Selection of hard constraints screen 1



(e) Selection of hard constraints screen 2



(f) Setting of optimization parameters

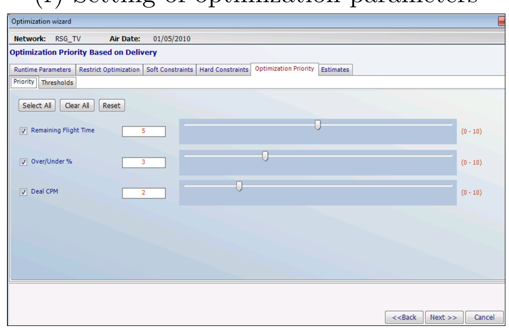


Figure A.3: Some screens of decision support system that package scheduling model.

Note. Screen shots of the application user interface (UI). The UI allows the user to define instances, select hard constraints, set soft constraint parameters, and set other optimization parameters.

Appendix B

Peer Effects In The Diffusion Of Solar Panels

B.1 Supporting Figures

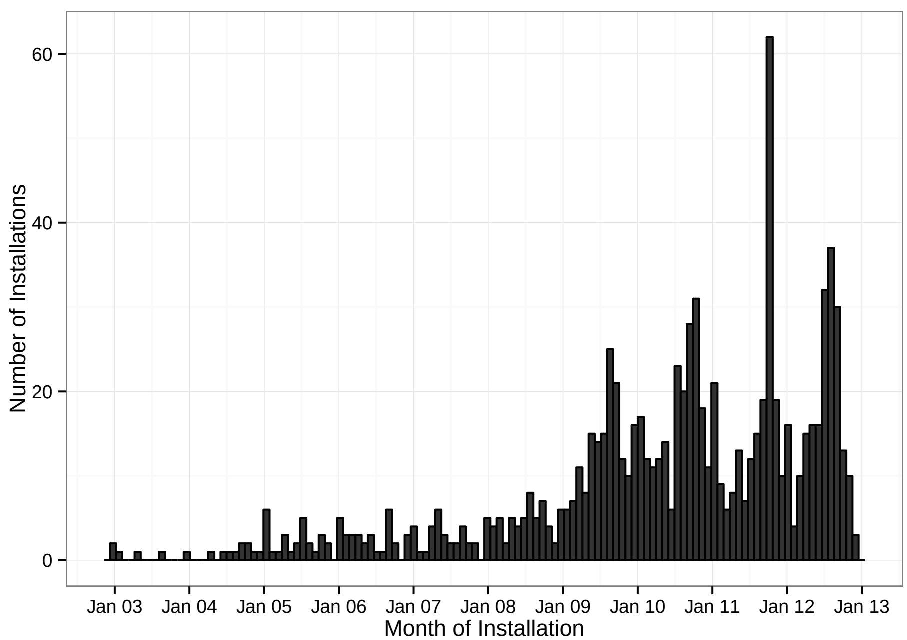
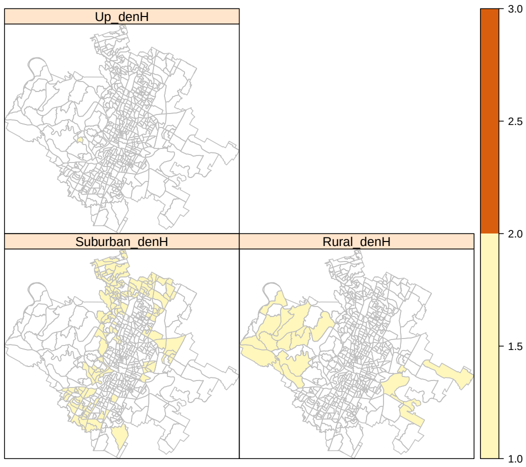
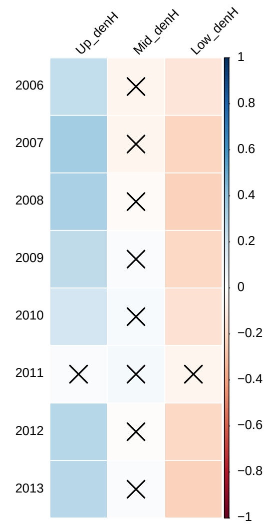


Figure B.1: Histogram of number of adopters per month in Austin between 2013 and 2013.
Source: Austin Energy.

(a) N households of households Income segment/N households per block group.



(b) Correlation adopters per year and households Income segment density. × indicates no significant.



(c) Dominant households Income segment per block group.

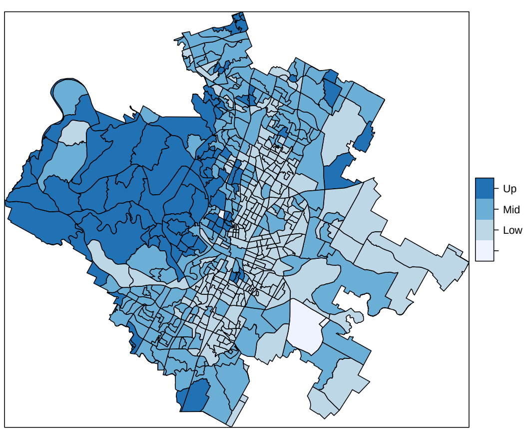
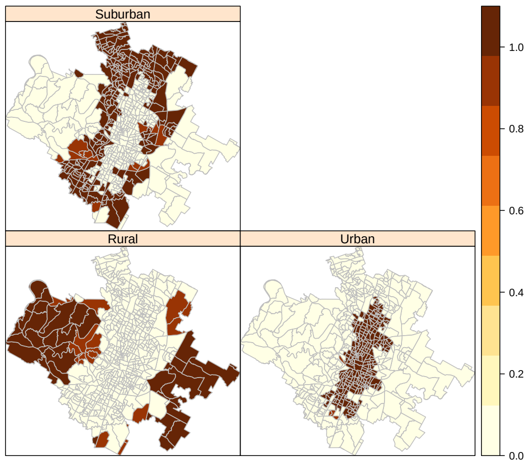
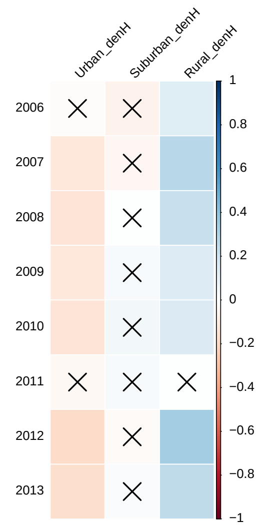


Figure B.2: Households income segmentation.
Source: Nielsen PRIZM.

(a) N households of Geography/N households per block group.



(b) Correlation adopters per year and geography density. × indicates no significant.



(c) Dominant Geography per block group.

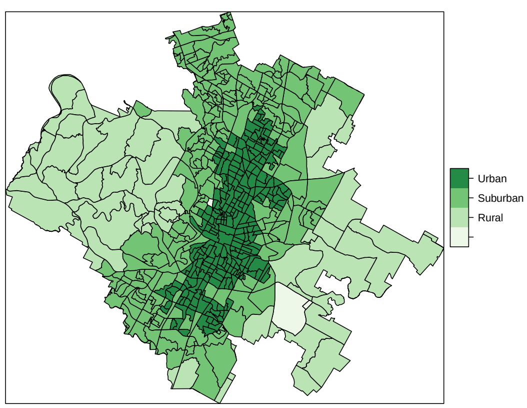
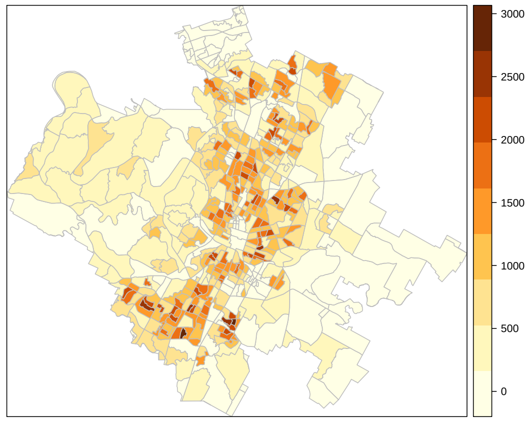


Figure B.3: Households geographic segmentation.
Source: Claritas PRIZM.

(a) N households / square mile per block group



(b) N adopters / N households per block group

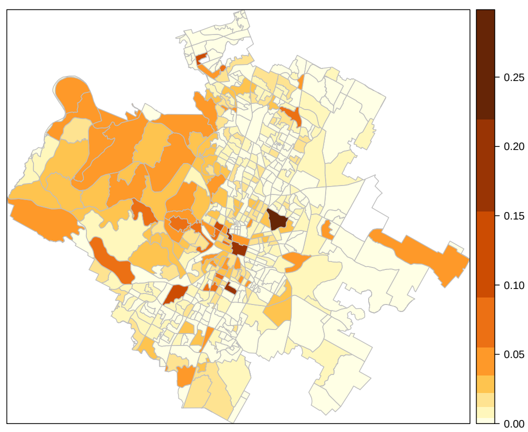
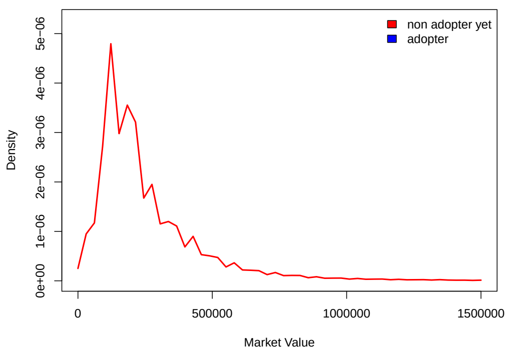
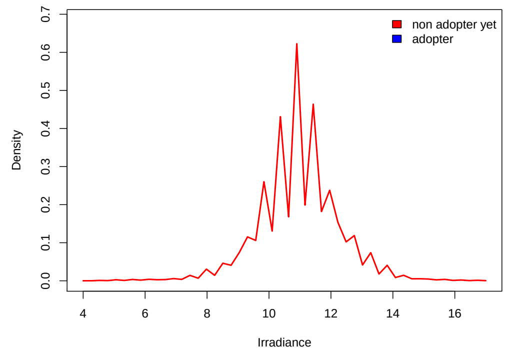


Figure B.4: Austin Energy service area and census block group division .
Source: Austin Energy.

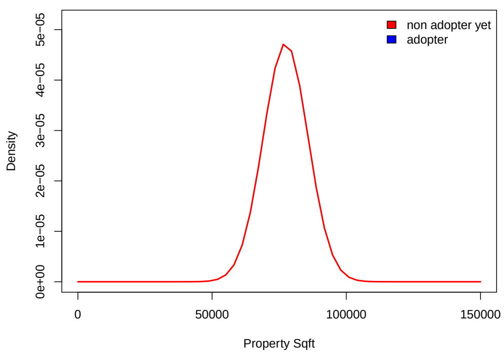
(a) Property Market Value.



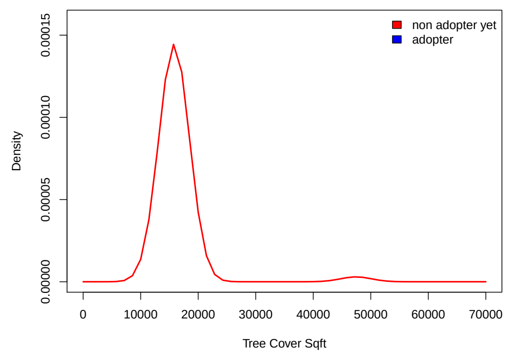
(b) Irradiance.



(c) Property Square Feet.



(d) Tree Cover Square Feet.



(e) Annual estimated system electricity generation.

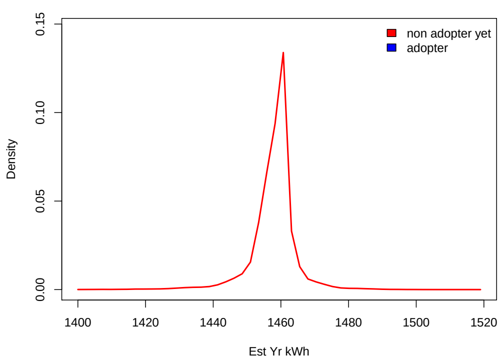


Figure B.5: Density of variables used to create clusters.

Bibliography

[1] Jaap H. Abbring and Øystein Daljord. Identifying the Discount Factor in Dynamic Discrete Choice Models. *Available at arXiv:1808.10651*, aug 2018.

[2] Victor Aguirregabiria and Pedro Mira. Dynamic discrete choice structural models: A survey. *Journal of Econometrics*, 156(1):38–67, may 2010.

[3] I Ajzen. The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2):179–211, 1991.

[4] V. F. Araman and I. Popescu. Media Revenue Management with Audience Uncertainty: Balancing Upfront and Spot Market Sales. *Manufacturing & Service Operations Management*, 12(2):190–212, June 2010.

[5] Arash Asadpour, Xuan Wang, and Jiawei Zhang. Online Resource Allocation with Limited Flexibility. 2016.

[6] M. Banciu, E. Gal-Or, and P. Mirchandani. Bundling Strategies When Products Are Vertically Differentiated and Capacities Are Limited. *Management Science*, 56(12): 2207–2223, December 2010.

[7] F.M Bass. A New Product Growth Model for Consumer Durable. *Management Science*, 15(January):215–227, 1969.

[8] FM Bass, TV Krishnan, and DC Jain. Why the Bass model fits without decision variables. *Marketing science*, 13(3):203–223, 1994.

[9] H. Blumenthal and O. R. Goodenough. *This business of television*. Billboard Books, 3 rev upd edition, 2006.

[10] S. Bollapragada and M. Garbiras. Scheduling Commercials on Broadcast Television, 2004.

[11] S. Bollapragada and S. Mallik. Managing on-air ad inventory in broadcast television. *IIE Transactions*, 40(12):1107–1123, oct 2008.

[12] S. Bollapragada, H. Cheng, M. Phillips, M. Garbiras, M. Scholes, T. Gibbs, and M. Humphreville. NBC's Optimization Systems Increase Revenues and Productivity. *Interfaces*, 32(1):47–60, January 2002.

[13] S. Bollapragada, M. R. Bussieck, and S. Mallik. Scheduling Commercial Videotapes in Broadcast Television, 2004.

[14] S. Bollapragada, S. Gupta, B. Hurwitz, P. Miles, and R. Tyagi. NBC-Universal uses a novel qualitative forecasting technique to predict advertising demand. *Interfaces*, 38 (2):103–111, March 2008.

[15] Bryan Bollinger and Kenneth Gillingham. Peer Effects in the Diffusion of Solar Photovoltaic Panels. *Marketing Science*, 31(6):900–912, sep 2012.

[16] M. J. Brusco. Scheduling advertising slots for television. *Journal of the Operational Research Society*, 59(10):1363–1372, aug 2008.

[17] M. J. Brusco and R. Singh. Assigning television commercial videotapes to time slots under alternative message spacing policies. *International Journal of Advertising*, 29: 431–450, 2010.

[18] José Antonio Carbajal and Wes Chaar. Turner Optimizes the Allocation of Audience Deficiency Units. *Interfaces*, 47(6):518 – 536, 2017.

[19] R. E. Caves and K. Guo. *Switching channels: Organization and change in TV broadcasting*. Harvard University Press, 2009.

[20] Deepa Chandrasekaran and Gerard J. Tellis. Diffusion and Growth of New Products : A Critical Review of Models and Findings. *Rev. Marketing*, pages 39–80, 2007.

[21] Renaud Chicoisne, Daniel Espinoza, Marcos Goycoolea, Eduardo Moreno, and Enrique Rubio. A New Algorithm for the Open-Pit Mine Production Scheduling Problem. *Operations Research*, 60(3):517–528, jun 2012.

[22] M. C. Chou, G. a. Chua, C.-P. Teo, and H. Zheng. Design for Process Flexibility: Efficiency of the Long Chain and Sparse Structure. *Operations Research*, 58(1):43–58, 2010.

[23] M. C. Chou, G. a. Chua, C.-P. Teo, and H. Zheng. Process Flexibility Revisited: The Graph Expander and Its Applications. *Operations Research*, 59(5):1090–1105, 2011.

[24] William S. Cleveland, Susan J. Devlin, and Eric Grosse. Regression by local fitting. Methods, properties, and computational algorithms. *Journal of Econometrics*, 37(1): 87–114, 1988. ISSN 03044076. doi: 10.1016/0304-4076(88)90077-2.

[25] P. Danaher and T. Dagger. Using a nested logit model to forecast television ratings. *International Journal of Forecasting*, 28:607–622, 2012.

[26] P. J. Danaher, T. S. Dagger, and M. S. Smith. Forecasting television ratings. *International Journal of Forecasting*, 27(4):1215–1240, October 2011.

[27] Yaniv Dover, Jacob Goldenberg, and Daniel Shapira. Network Traces on Penetration: Uncovering Degree Distribution from Adoption Data. *Marketing Science*, 31:689–712, 2012.

[28] D Espinoza, R Garcia, M Goycoolea, G L Nemhauser, and M W P Savelsbergh. Per-Seat, On-Demand Air Transportation Part II: Parallel Local Search. *Transportation Science*, 42(3):279–291, aug 2008.

[29] Tal Garber, Jacob Goldenberg, Barak Libai, and Eitan Muller. From Density to Destiny: Using Spatial Dimension of Sales Data for Early Prediction of New Product Success. *Marketing Science*, 23(3):419–428, jun 2004.

[30] D. R. Gaur, R. Krishnamurti, and R. Kohli. Conflict Resolution in the Scheduling of Television Commercials. *Operations Research*, 57(5):1098–1105, March 2009.

[31] Andrew Gelman, John B Carlin, Hal S Stern, and Donald B Rubin. Evaluating, comparing, and expanding models. In *Bayesian Data Analysis*, chapter 7, pages 165–195. CRC Press, third edition, 2004.

[32] Jacob Goldenberg, Sangman Han, Donald R Lehmann, and Jae Weon Hong. The Role of Hubs in the Adoption Process. *Journal of Marketing*, 73:1–13, 2009.

[33] Jacob Goldenberg, Barak Libai, and Eitan Muller. The chilling effects of network externalities. *International Journal of Research in Marketing*, 27:4–15, 2010.

[34] Stephen C. Graves and Brian T. Tomlin. Process Flexibility in Supply Chains. *Management Science*, 49(7):907–919, 2003.

[35] Marcello Graziano and Kenneth Gillingham. Spatial patterns of solar photovoltaic system adoption: the influence of neighbors and the built environment. *Journal of Economic Geography*, oct 2014.

[36] A. Grosfeld-Nir and Y. Gerchak. Multiple lotsizing in production to order with random yields: Review of recent advances. *Annals of Operations Research*, 126:43–69, 2004.

[37] R. S. Headen, J. E. Klompmaker, and J. E. Teel. Predicting Audience Exposure to Spot TV Advertising Schedules. *Journal of Marketing Research*, 14:1–9, 1977.

[38] Yansong Hu and Christophe Van den Bulte. Nonmonotonic Status Effects in New Product Adoption. *Marketing Science*, (Forthcoming), 2014.

[39] Rob J Hyndman. *forecast: Forecasting functions for time series and linear models*, 2016. R package version 7.3.

[40] Susumu Imai, Neelam Jain, and Andrew Ching. Bayesian Estimation of Dynamic Discrete Choice Models. *Econometrica*, 77(6):1865–1899, nov 2009.

[41] R. Iyengar, C. Van den Bulte, and T. W. Valente. Opinion Leadership and Social Contagion in New Product Diffusion. *Marketing Science*, 30(2):195–212, dec 2011.

[42] W. C. Jordan and S. C. Graves. Principles on the Benefits of Manufacturing Process Flexibility. *Management Science*, 41(4):577–594, 1995.

[43] Elmar Kiesling, Markus Guenther, Christian Stummer, and Lea M Wakolbinger. Agent-based simulation of innovation diffusion: a review. *Central European Journal of Operations Research*, 20:183–230, 2012.

[44] A. Kimms and M. Muller-Bungart. Revenue management for broadcasting commercials: the channel's problem of selecting and scheduling the advertisements to be aired. *Internat. Journal of Revenue Management*, 1(1):28–44, 2007.

[45] Ruben Lobel and Georgia Perakis. Consumer Choice Model For Forecasting Demand And Designing Incentives For Solar Technology. 2013.

[46] Charles F. Manski. Identification of Social Endogenous Effects: The Reflection Problem. *The Review of Economic Studies*, 60(3):531–542, 1993.

[47] Charles F Manski and Steven R Lerman. The Estimation of Choice Probabilities from Choice Based Samples. *Econometrica*, 45(8):1977–1988, nov 1977.

[48] Andriy Norets. Inference in Dynamic Discrete Choice Models With Serially Correlated Unobserved State Variables. *Econometrica*, 77(5):1665–1682, 2009.

[49] J. Palmer, G. Sorda, and R. Madlener. Modeling the diffusion of residential photovoltaic systems in Italy: An agent-based simulation. *Technological Forecasting and Social Change*, 99:106–131, oct 2015.

[50] Shinjini Pandey, Goutam Dutta, and Harit Joshi. Survey on revenue management in media and broadcasting. *Interfaces*, 47(3):195–213, 2017.

[51] Renana Peres, Eitan Muller, and Vijay Mahajan. Innovation diffusion and new product growth models: A critical review and research directions. *International Journal of Research in Marketing*, 27:91–106, 2010.

[52] R. Phillips and G. Young. Television Advertisement Pricing in the U.S. In Özalp Özer Phillips and Robert, editors, *Oxford Handbook of Pricing Management*, chapter 14, pages 230–249. Oxford University Press, New York, 2012.

[53] Dana G. Popescu and Pascale Crama. Ad Revenue Optimization in Live Broadcasting. *Management Science*, 62(4):1145–1164, 2016.

[54] Hazhir Rahmandad and John Sterman. Heterogeneity and Network Structure in the Dynamics of Diffusion: Comparing Agent-Based and Differential Equation Models. *Management Science*, 54(5):998–1014, 2008.

[55] Varun Rai and Kristine McAndrews. Decision-making and behavior change in residential adopters of solar PV. In *World Renewable Energy Forum (Denver, CO)*, 2012.

[56] Varun Rai and Scott Robinson. Effective information channels for reducing costs of environmentally- friendly technologies: evidence from residential PV markets. *Environmental Research Letters*, 8(1):014044, mar 2013.

[57] Varun Rai and Scott A. Robinson. Agent-based modeling of energy technology adoption: Empirical integration of social, behavioral, economic, and environmental factors. *Environmental Modelling & Software*, 70:163–177, aug 2015.

[58] S. K. Reddy, J. E. Aronson, and A. Stam. SPOT: Scheduling Programs Optimally for Television. *Management Science*, 44(1):83–102, January 1998.

[59] G. O. Roberts, A. Gelman, and W. R. Gilks. Weak convergence and optimal scaling of random walk Metropolis algorithms. *Annals of Applied Probability*, 7(1):110–120, 1997.

[60] E. Rogers. *Diffusion of Innovations, 5th Edition*. Free Press, first edition, 2003.

[61] Everett M Rogers. *Diffusion of Innovations*. Free Press, first edition, 1962.

[62] John Rust. Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica: Journal of the Econometric Society*, 55(5):999–1033, 1987.

[63] R. T. Rust, W. A. Kamakura, and M. I. Alpert. Viewer Preference Segmentation and Viewing Choice Models for Network Television. *Journal of Advertising*, 21(1):1–18, 1992.

[64] Andrzej Ruszczyński and Alexander Shapiro. Two- Stage Problems. In Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczyński, editors, *Lectures on Stochastic Programming: Modeling and Theory*, chapter 2, pages 27–62. SIAM, 1st edition, 2009.

[65] Nazrul I. Shaikh, Arvind Rangaswamy, and Anant Balakrishnan. Modeling the Diffusion of Innovations Through Small-World Networks. *Available at SSRN*, oct 2010. ISSN 1556-5068. doi: 10.2139/ssrn.2032861. URL http://www.ssrn.com/abstract=2032861.

[66] R. H. Shumway and D. S. Stoffer. *Time Series Analysis and Its Applications*. Springer, third edition, 2011. ISBN 9781441978646.

[67] D. Simchi-Levi and Y. Wei. Understanding the Performance of the Long Chain and Sparse Designs in Process Flexibility. *Operations Research*, 60(5):1125–1141, 2012.

[68] David Simchi-Levi and Yehua Wei. Worst-Case Analysis of Process Flexibility Designs. *Operations Research*, 63(1):166–185, 2015.

[69] J. Smith and C. Ulu. Risk Aversion, Information Acquisition, and Technology Adoption. *Operations Research*, 65(4):1011–1028, 2017.

[70] James E. Smith and Canan Ulu. Technology Adoption with Uncertain Future Costs and Quality. *Operations Research*, 60(2):262–274, may 2012.

[71] Ashish Sood, Gareth M James, and Gerard J Tellis. Functional Regression: A New Model for Predicting Market Penetration of New Products. *Marketing Science*, 28: 36–51, 2009.

[72] Ashish Sood, Gareth M James, Gerard J Tellis, and Ji Zhu. Predicting the Path of Technological Innovation: SAW vs. Moore, Bass, Gompertz, and Kryder. *Marketing Science*, 31(6):964–979, 2012.

[73] David J. Spiegelhalter, Nicola G. Best, Bradley P. Carlin, and Angelika van der Linde. Bayesian Measures of Model Complexity anf Fit. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, 64(4):583–639, 2002.

[74] J. Stachurski. *Economic Dynamics: Theory and Computation*. The MIT Press, 2009.

[75] K. T. Talluri and G. J. Van Ryzin. Media and Broadcating. In *The theory and practice of revenue management*, chapter 10.5, pages 542–546. Springer, 2005.

[76] Canan Ulu and James E. Smith. Uncertainty, Information Acquisition, and Technology Adoption. *Operations Research*, 57(3):740–752, mar 2009.

[77] C Van den Bulte and S Stremersch. Social contagion and income heterogeneity in new product diffusion: A meta-analytic test. *Marketing Science*, 23:530–544, 2004.

[78] Christophe Van den Bulte. New Product Diffusion Acceleration: Measurement and Analysis. *Marketing Science*, 19(4):366–380, nov 2000.

[79] Christophe Van den Bulte and Yogesh V. Joshi. New Product Diffusion with Influentials and Imitators. *Marketing Science*, 26(3):400–421, may 2007.

[80] Xuan Wang and Jiawei Zhang. Process Flexibility: A Distribution-Free Bound on the Performance of k -Chain. *Operations Research*, 63(3):555–571, 2015.

[81] J. Webster, P. Phalen, and L. Lichty. *Ratings Analysis: Audience Measurement and Analytics*. Routledge, 4th editio edition, October 2013. ISBN 080585410X.

[82] G. Y. Weintraub, C. L. Benkard, and B. Van Roy. Markov Perfect Industry Dynamics With Many Firms. *Econometrica*, 76(6):1375–1411, 2008.

[83] G. Y. Weintraub, C. L. Benkard, and B. Van Roy. Computational Methods for Oblivious Equilibrium. *Operations Research*, 58(4-Part-2):1247–1265, jun 2010.

[84] K. Wilbur, M. Goeree, and G. Ridder. Effects of advertising and product placement on television audiences. *Available at SSRN 1151507*, 2008.

[85] X. Zhang. Mathematical models for the television advertising allocation problem. *International Journal of Operational Research*, 1(3):302–322, 2006.

Vita

Sebastián A. Souyris Ilabaca was born in Santiago, Chile. He received his B.A. in Civil Industrial Engineering and M.S. in Operations Management (OM) from the University of Chile, then completed his M.Phil. in OM at NYU Stern School of Business. In 2019, he finished his Ph.D. in OM at the McCombs School of Business at The University of Texas Austin, and is joining the Gies College of Business at the University of Illinois at Urbana-Champaign as a Postdoctoral Research Associate. At the University of Chile he taught optimization and OM courses in the Department of Industrial Engineering. In Santiago, Sebastián worked as an assistant professor at the Faculty of Engineering at the University of Los Andes. Recently in NYC, he served as the Vice President of Data Science for RSG Media, developing analytical methods for TV networks.

Permanent address: sebastian.souyris@utexas.edu

This dissertation was typeset by the author.